University of Alberta

Collapse-Time Distribution for Large Cosmic Structures

by

Kipp Cannon ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial
fulfillment of the requirements for the degree of Doctor of Philosophy

Department of Physics

Edmonton, Alberta
Spring 2004

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-612-96247-4
Our file Notre référence
ISBN: 0-612-96247-4

The author has granted a non-exclusive license allowing the Library and Archives Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou aturement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

This thesis presents a computation of the collapse-time distribution for large-scale structures in the universe as well as a determination of the time evolution of that distribution given a precise, instantaneous, measurement of an horizon-limited volume of the universe. The presentation begins with a simple review of the generation of cosmological perturbations during the inflation epoch. The description of the perturbations is carried out to linear-order using the conformal Newtonian (zero-shear) gauge and assuming a $\frac{1}{2}m^2\phi^2$ potential for the inflaton field. A fully non-linear model for the collapse of the cosmological fluid to a singularity is then developed using the Tolman-Lemaître line-element. This collapse model assumes that the cosmological fluid is irrotational and that the time required for the collapse of the material is a monotonically increasing function of proper radius. Matching conditions are determined, allowing for the construction of initial conditions for the Tolman-Lemaître collapse model from the linear Newtonian perturbations. Using the collapse model and current cosmological data, a distribution is obtained for the proper collapse time of a shell of material as a function of radius. An observer, located at the centre of the shells, is then considered and is imagined to be able to determine exactly the collapse time of the material within some horizon-limited volume of the universe. The time evolution of the collapse-time distribution for the visible volume of the universe is then determined based on the increasing radius of the horizon and the observation. It is found that the most concise expression of the result of this analysis is the median of the distribution of proper collapse times for the visible volume of the universe as a function of elapsed proper time for the centrally-located observer. This function is found to have a near power-law behaviour at early and late times, although its algebraic structure is somewhat complex.

Table of Contents

1 Introduction **1**
  1.1 Road Map 5

2 Cosmological Inflation **8**
  2.1 Friedmann-Robertson-Walker Model 8
  2.2 Behaviour Of Scalar Fields 13
  2.3 Application To Cosmology 14

3 The Description of Cosmological Perturbations **22**
  3.1 Gaussian Random Fields 22
  3.2 Random Fields and Cosmology 29
    3.2.1 Power Spectrum 30
    3.2.2 Spectral Index 33
  3.3 Units 34
  3.4 Properties of Gaussian Random Fields 35
    3.4.1 Smoothing 35
    3.4.2 The Properties of $\overline{f}$, Part I 41
    3.4.3 The Properties of $\tilde{f}$ 44
    3.4.4 The Properties of $\overline{f}$, Part II 47
  3.5 Matter Distribution 52

4 The Behaviour of Cosmological Perturbations **53**
  4.1 Quantum Inflaton Perturbations 54
  4.2 The Transition to Classical Perturbations 58
    4.2.1 The "Curvature Perturbation" 60
  4.3 Geometry Perturbations — Description 61
  4.4 Geometry Perturbations — Behaviour 66
    4.4.1 Refinement of Inflaton Perturbations 70
  4.5 Comparison with Observation 71
    4.5.1 Fluctuation Amplitude 71

5 Collapse **74**
  5.1 Spherically-Symmetric Model 74
    5.1.1 Overview 74

5.1.2 Tolman-Lemaître Geometry 77
5.1.3 Construction of $t_{\text{collapse}}(r)$ From Invariant Scalar Expressions . . . 84
5.1.4 Construction of $t_{\text{collapse}}(r)$ By Geometry Matching 89
5.1.5 Approximating the Velocity Perturbations 96
5.1.6 Evaluating t_{collapse} 97
5.2 The Collapse-Time Distribution 98
5.2.1 The Distribution of $\frac{1}{aH}\Phi'(r)$ 98
5.2.2 The Distribution of $t_{\text{collapse}}(r)$ 103
5.2.3 Commentary 110
5.3 Incorporating a Cosmological Constant 112
5.3.1 Dust Condensing out of an Inflating Background 112
5.3.2 Commentary 116
5.4 The Evolution of the Collapse-Time Distribution 117
5.4.1 The Radial Two-Point Joint Collapse-Time Distribution 120
5.4.2 Commentary 127
5.5 Comparison to the Press-Schechter Model 132
5.5.1 Overview of the Press-Schechter Model 132
5.5.2 Connection to the Formalism of this Document 133
5.5.3 Collapse-Time for a Spherical Top-Hat Fluctuation 134

6 Conclusions **136**
6.1 Summary 136
6.2 Recent Revisions of Cosmological Parameters 137
6.3 Possible Extensions to this Work 138

Bibliography **141**

A Units and Sign Conventions **149**

B Numerical Parameters **151**

C Derivations **154**
C.1 Solving Einstein's Field Equation for the Geometry of Space-Time 154
C.2 Christoffel Symbols for the RW Line Element 155
C.3 FRW Mass Continuity Equation in (2.1.7) 156
C.4 Equation (2.1.9) 157
C.5 Equation of Motion for the Scalar Field in (2.2.1) 157
C.6 The Removal of the Phase Factor in (3.4.40) 159
C.7 Solution to (4.1.13) 160
C.7.1 Notes on Bessel Functions 164
C.8 The Derivation of (4.1.18) 165
C.9 Linear Cosmological Perturbations 166
C.9.1 Equations of Motion for the Newtonian Potential and Inflaton Perturbations 166

C.9.2 Equations of Motion for the Newtonian Potential and Dust Perturbations . . . 175
C.9.3 Spherical Co-ordinates . . . 179
C.10 The Derivation of (4.4.7) . . . 180
C.11 Notes on Hypergeometric Functions . . . 181
C.12 The Derivation of (5.2.14) . . . 184
C.13 The Derivation of (5.4.10) . . . 186

D Random Variables **187**
D.1 Propagation of Distribution Densities . . . 187
D.2 Gaussian Random Variables and Processes . . . 189
D.2.1 The Distribution of $z = ax$ . . . 189
D.2.2 The Distribution of $z = x + y$ . . . 190
D.3 The Dirac δ . . . 191

List of Tables

3.1 Some definitions of "power spectrum" found in the wild in terms of the equivalent quantities used in this document. 31
3.2 The units carried by the various quantities used to describe random fields in terms of the units carried by the co-ordinates and the field itself. 35

4.1 Relationships between the perturbation variables used in this document and those of [6] and [69]. 64

6.1 Revised cosmological parameters from the WMAP. 138

B.1 Physical constants and conversion factors. 151
B.2 Derived physical constants and conversion factors. 152
B.3 Additional conversion factors. 152
B.4 Conversion factors derived from cosmological parameters. 152
B.5 Cosmological parameters. 153

List of Figures

1.1 Aristotle's proof that we must be at the centre of the universe. 2
1.2 Determining the arrangement of the Sun, Earth and Moon by finding two of the angles in the triangle. 2

2.1 Numerical integration of (2.3.11) and (2.3.12) showing the typical time dependence of ϕ and H during and immediately following the inflation epoch. 19

3.1 The distribution of the modulus and argument of a single mode with a spectral density of $S_{\vec{k}} = 1$ in the Fourier transform of a homogeneous Gaussian random field. 28
3.2 The radial behaviour of the two standard window functions. 36
3.3 Smoothing a realization of a random field in one dimension. 38
3.4 The frequency dependence of the two standard window functions. 41
3.5 The behaviour of $\left\langle \overline{f}^2 \right\rangle \left(\sigma^2 / r^{n_s - 1}\right)^{-1}$ as a function of the spectral index n_s for two choices of window function. 43
3.6 The spectral behaviour of $W_{\vec{k};r} \left(1 - \sqrt{2\pi}^3 W_{\vec{k};r_*}\right)$ for $r_*/r = 10$ illustrating the long wavelength filtering. 48
3.7 The behaviour of $\left\langle \overline{f}^2 \right\rangle \left(\sigma^2 / r^{n_s - 1}\right)^{-1}$ as a function of the spectral index n_s for four choices of $\frac{r_*}{r}$ using Gaussian windows. 51
3.8 The behaviour of $\left\langle \overline{f}^2 \right\rangle \left(\sigma^2 / r^{n_s - 1}\right)^{-1}$ as a function of $\frac{r_*}{r}$ for a fluctuation spectrum with $n_s = 1$ illustrating the logarithmic divergence of the mean square with increasing r_*. 51

5.1 The schematic relationship between conformal Newtonian co-ordinates and Tolman-Lemaître co-ordinates at the space-time event $\mathcal{A}$. 91
5.2 The mean square of $\frac{1}{aH}\Phi'(raH)$ after smoothing $n_{s\Phi} = 1$ perturbations using two techniques for choosing the smoothing scale. 102
5.3 The mean square of $\frac{1}{aH}\Phi'(raH)$ as a function of $n_{s\Phi}$ for several radii with $r_0 aH = 1$. 103
5.4 The relationship between t_{collapse} and $\tau^{-\frac{2}{3}}$. 106
5.5 Cumulative probability of collapse $P\left(\text{collapse} \in [\tau^{-\frac{2}{3}}, \infty)\right)$ as a function of raH for a spectral index of $n_{s\Phi} = 1.01$ and smoothing radius of $r_0 aH = 1$. 108

5.6 The cumulative probability of collapse, $P\left(\text{collapse} \in [\tau^{-\frac{2}{3}}, \infty)\right)$, at the smoothing radius, $raH = r_0aH = 1$, as it depends on the spectral index, $n_{\mathrm{s}\Phi}$. . . . 109
5.7 The median reduced inverse collapse time as a function of raH/r_0aH for several choices of smoothing radius, and a range of spectral indices at each. 110
5.8 The probability of condensation as a function of region size for matter density spectral indices in the range $4 \leq n_{\mathrm{s}\delta} \leq 6$, corresponding to $0 \leq n_{\mathrm{s}\mathcal{R}} \leq 2$. . . . 114
5.9 The probability of condensation as a function of region size for $n_{\mathrm{s}\delta} = 5.01^{0.09}_{0.08}$. 115
5.10 The radius to which an observer can see as a function of time. . . . 119
5.11 The increase of horizon radius as a function of elapsed time. . . . 120
5.12 The cumulative probability of collapse $P\left(\text{collapse} \in [\tau^{-\frac{2}{3}}(r), \infty)\right)$, conditional on the material inside r_1aH being critical, as it depends on radius. 123
5.13 A plot of the median reduced inverse collapse time as a function of horizon radius using the current best-estimate for the spectral index. . . . 124
5.14 A plot of the median reduced inverse collapse time as a function of elapsed proper time for a centrally-located observer using the current best estimate for the spectral index. . . . 125
5.15 A plot of the median reduced inverse collapse time as a function of elapsed proper time for a centrally-located observer for several spectral indices. . . 126
5.16 The effect of reducing the smoothing radius on the behaviour of the median reduced inverse collapse time. . . . 127
5.17 The effect of changing the radius of initial observation on the behaviour of the median reduced inverse collapse time. . . . 128
5.18 A repeat of Figure 5.14 with both axes in Planck times. . . . 129
5.19 A repeat of Figure 5.18 with a polynomial fit at small times. . . . 132
5.20 The collapse time Ht_{collapse} (in Hubble times) for a spherical top-hat fluctuation of initial proper radius r_iaH (in Hubble radii) in an expanding, critical-density, background normalized to the fluctuation's initial density contrast, δ_i. . . . 135

C.1 The result of the integral in (C.6.1) as a function of the upper bound, $k_{\max}$, for a Gaussian window. . . . 159

Chapter 1

Introduction

Throughout history, humanity's view of how special we are in the grand scheme of the cosmos has varied. Until the Copernican revolution, the most common view had been that we were located at a very unique place.[1] Although the "we-are-the-centre-of-the-universe" view might cause people to chuckle today, it is not actually unwarranted given the limited data available to naked-eye astronomers.

Consider the stars, for example. The stars are clearly among the farthest objects in the sky since clouds, birds, etc., can all obscure them and yet to the unaided eye the stars do not appear to move. The two possible conclusions are that either the stars are so far away that any hypothetical motion of the Earth cannot reveal itself in their positions, or that the Earth's location is fixed with respect to them. Not only is the latter more plausible to the ancient mind, considering the phenomenal brightness with which the stars would have to shine in order to be visible should the former be the case, but the latter option also has the attraction of providing more information about the structure of the cosmos. An example of the sort of conclusion that can be extracted from the "close stars" model of the cosmos is shown in Figure 1.1. This is a sixteenth-century woodcut illustrating the reasoning put forward by Aristotle for deducing our location in the cosmos. The argument is quite simple: since, whenever we look up, we always see exactly half of the heavens, we must be at the centre. This argument breaks down if the stars are at a very great distance from us.

Now consider the moon. All one needs to do is assume that the Moon is a sphere and it quickly becomes clear from correlating the Moon's crescent with the apparent position of the Sun that the Sun is the source of the Moon's illumination. The position of the day-night terminator on the Moon's surface then indicates the angle between our line of sight to the Moon and the Moon's line of sight to the Sun. Combined with the angle between our lines of sight to the Sun and the Moon, it is easy to draw the triangle that is the Sun-Earth-Moon system. The arrangement is shown in Figure 1.2. The scale of the triangle cannot be known without more accurate measurements but that is not important. What is important is that by repeating these measurements and drawing the shape of the triangle throughout the month it becomes apparent that the Moon is going

---

[1] See, for example, [9, Chapter 1].

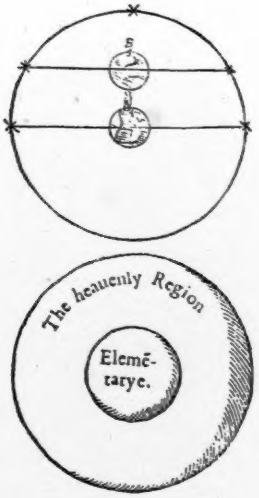


Figure 1.1: A sixteenth-century wood-cut illustrating Aristotle's proof that we must be at the centre of the universe: when we look up, we never see less than half the sky. Taken from [35].

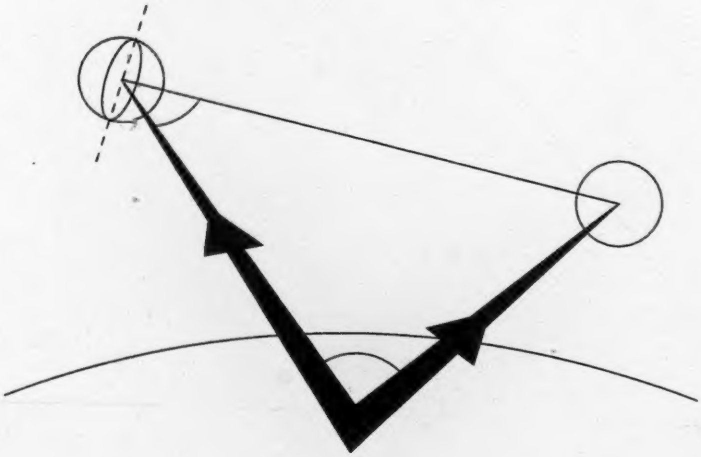


Figure 1.2: Determining the arrangement of the Sun, Earth and Moon by finding two of the angles in the triangle. One angle is found by direct measurement, the other by observing the apparent location of the Moon's day-night terminator.

around the Earth, not the Sun. This only makes sense: if, as revealed by the stars, we are at the centre of the cosmos then anything that's moving must be moving around us.

As ever more precise data was collected on the motions of the planets, the Ptolemaic model of the cosmos, with us as the centre of it all, could be refined farther and farther. Toward the end of the model's life it had become quite complex but most interestingly, and contrary to popular understanding, the Earth was no longer at the centre. The centre of each of the circles along which the planets travelled was assigned a time-dependant offset in order to allow the model to more closely (and easily) match observation. Although the model still placed the Earth near the centre of the cosmos, data had superseded philosophy and placed the Earth a little ways away from it. In fact, as early as the third century BC, models of the cosmos had been proposed in which Mercury and Venus were placed in orbits around the Sun[2]. These models were motivated purely by the observation that Mercury and Venus are always seen to be close to the Sun in the sky.

When people, most famously Galileo Galilei, first pointed telescopes at the sky startling new data became available. Galileo Galilei in particular observed several remarkable phenomena.[3] Firstly he observed that in increasing the magnification of his telescopes, the apparent diameter of each of the planets would increase, however the stars continued to appear to be just points of light. This made it much easier to believe that the stars are perhaps at a very great distance from the Earth — potentially so great to be of no use in deducing the Earth's position or motion through parallax measurements. He also observed "stars" orbiting Jupiter. It was clear that they were associated with Jupiter since they moved with it from one night to the next relative to the background stars. Finally, he also saw Venus exhibit phases like those of the moon. When one uses the phases, as above, to repeatedly plot the Sun-Earth-Venus triangle one finds that Venus is not orbiting the Earth at all; it is, in fact, orbiting the Sun.

The telescope, then, revealed that many objects are not orbiting the Earth, and in fact of all objects about which there was then specific information the only object not orbiting something other than Earth was the Moon. Our place in the cosmos began to look decidedly less special.

More recently and on a grander scale, surveys of galaxy positions on the sky reveal that they are isotropically distributed.[4] In the absence of precise range information, just enough to establish that galaxies are distant objects, one of two conclusions is possible. The first is that we are sitting at the one special location in the universe where the distribution of galaxies happens to look isotropic: the centre of a spherical, layered, onion-like structure. The second possibility is that we are not at a special location; that the galaxies are homogeneously distributed throughout the visible universe so that their distribution appears isotropic from all vantage points, including our own. Unlike the philosophical leanings of their predecessors, it is the second possibility, that we are in no way special, that has generally been favoured by modern cosmologists.

This means that when attempting to come to an understanding of newly-observed or predicted phenomena, it is often assumed that the investigation can only reveal that our observations are as normal as possible.[5] Observations don't always make this an easy position to take. Using the tool of general relativity to analyze the universe's life cycle reveals that our visible universe is not "uninteresting" at all but is actually extremely special. Until recently, our understanding of physics did not provide any reason for the initial conditions of the universe to have had any particular values. In the classical big bang model, however, if the initial density of the universe differed from the universe's critical density — its gravitational balance point — by more than just one part in 10^{55} within a Planck-time of its creation, the visible universe could not have grown as old as it has while remaining as dense as it is.[6] If it had been more dense then it would have collapsed too quickly and if less dense it would have dispersed too rapidly. Differing by less than one part in 10^{55} from the only special value is a rather striking co-incidence and has led researchers to look for reasons why the universe must have started the way it did. The philosophical desire for our universe to be typical inspires the search for the reasons why it must be so.

---

[2] These models are often ascribed to Heracleides of Pontus (c. 390 – c. 310 BC) [35, Chapter 4].

[3] See [35, Chapter 5].

[4] The original report of this phenomenon is in [36].

There is, however, the other possibility: perhaps the universe *is* special. This idea has received some serious attention as the "anthropic principle." Briefly, the idea is that humans are not simply impartial observers of the universe but our existence, in fact, plays a role in defining it. More precisely, consider that the universe was free to have its initial parameters take on any range of values as described by some distribution. The distribution might be broad or it might be very narrow depending on how much the fundamental physics constrains the outcome but whatever it is we don't know it and let's assume it's not a δ-function. Consider also the distribution representing the likelihood of humanity's eventual existence as a function of the early conditions of a universe. When one asks the question "how likely is the universe to be the way we see it," one is asking for the product of these two distributions, not simply the first.

The anthropic principle is the statement that if the distribution of conditions acceptable for our existence is sufficiently narrow then, regardless of the inherent distribution of outcomes expected from the fundamental physics, the universe as we see it will always be as expected. That's not to say that the distribution conditional on our existence need become a δ-function. It's fine if the resulting distribution allows for a range of possibilities, even including those that are quite different from what we see. The anthropic principle may still be taken as an explanation of our observations as long as within the likely range of outcomes our own universe is no longer absurdly special in any way.

An analogy can be made with our planetary environment. One might wonder how likely it is for us humans to have found ourselves living on a planet with abundant liquid water at the surface. It would be inappropriate, although perhaps tempting, to determine this probability by counting up all the planets in the universe with liquid water on their surface and dividing by the total (assuming, for the moment, that such numbers are available). The problem is that we have evolved as a species requiring abundant liquid water on a planet where that was the case so the probability of *us* having found ourselves on a such a planet is nearly 1 no matter how unusual such planets are. Likewise, investigating why it is that we are lucky enough to have found ourselves living in a universe that looks the way ours does may turn out to be equally uninteresting — the coincidence might be a necessity.

---

[5] See, for example, [8, Sections IIIA and IIIB].

[6] See Section 2.1.

There are two genuine difficulties with the anthropic principle. The first is that it is very important not to use it to avoid the proper investigation of phenomena. The anthropic principle does not explain how the Earth came to have so much water or how our species developed, it only says that there was no luck in the two processes coinciding. One certainly should not be deterred from attempting to understand them. The other problem with the anthropic principle is that it finds itself on very thin ice unless there are lots of chances for things to occur: there are, presumably, lots of planets out there and we simply formed on one capable of supporting our style of organism; we only know of the one universe, however, so in the absence of a mechanism by which others could have formed one is once again left saying "we were just that lucky."

In an attempt to explain the early features of our universe, to provide a reason for why they must have been the way they were, starting in the early 1980s researches developed what have come to be known as inflationary cosmological models. In these models the universe undergoes a period of accelerating growth which causes it to become very smooth and homogeneous — the requisite properties. One finds, however, that this growth leaves the universe much much larger than what we can observe today and so, ironically, cosmological inflation also provides a mechanism by which the early characteristics need not have been anything in particular at all. By providing a final scale so huge, inflation leaves us with many attempts at a suitable universe where each attempt is a distinct visible-universe-sized patch left over after the inflation epoch. The physics of the inflation epoch determines the distribution of initial conditions seen in the patches so one can quantitatively discuss how normal our universe is, or equivalently how different it could have been, in the context of various inflation models.

This document will present the calculation of the distribution of one characteristic of these post-inflationary patches: the distribution of their collapse times. Although inspired by the ideas of the anthropic principle, this computation does not impact on it or the conclusions one draws from it in any way. Since the apparent time it will take our visible universe to collapse is indistinguishable from infinity and this state is readily achieved as a result of cosmological inflation, there is no need for cosmological models to be constrained with regard to the collapse-time parameter; and in any case there can really be no observations with which to constrain them since we will not be around long enough to make them. The computation is, rather, an exercise in curiosity.

1.1 Road Map

Let me first explain the layout of this document. Overall, there are two parts. The first few chapters contain background material, or material that lays the ground-work for the computations to follow. These chapters, themselves, contain varying degrees of detail with much additional material hidden in appendices. Following these chapters is the work that addresses the specific topic of this document: cosmological collapse. The particular content of each chapter is as follows. The text will begin with a review in Chapter 2 of cosmological inflation. Following this, Chapter 3 is devoted to setting up the notation and relations to be used in describing random fields in general, and cosmological perturbations in particular. A review of cosmological perturbation theory is then presented in Chapter 4 with a description of the formation of super-horizon-size curvature perturbations during the inflation epoch. Finally, Chapter 5 consists of the development and analysis of several collapse models.

To be precise, the "cosmic structures" considered in this document are comoving spherical volumes of the universe centred on a comoving observer, where the spherical volumes have radii on the order of or larger than the Hubble radius. The time evolution of the material within such a spherical volume is analyzed on the condition that it collapses to a singularity. The "collapse time" of such a spherical volume is then defined to be the proper time that elapses along the centrally-located comoving observer's world line to the event of that world line's intersection with the collapsed singularity. The ensemble of choices of centrally-located observers is used to construct an ensemble of collapse histories from which the statistical properties of the collapse process can be analyzed. What shall be obtained in particular is the way in which the distribution for the collapse time evolves over time given a single horizon-limited observation of the universe by the centrally-located observer.

Let me now give an overview of the computation by which the distribution of collapse times will be obtained. What will be done in the chapters to follow will be to begin by establishing a model for determining the collapse time for a spherical region of space. The spectrum of curvature fluctuations expected to be present following a period of cosmological inflation will then be used to establish a distribution for the initial conditions of such a spherical region and from that obtain a distribution for the collapse time. As a second step, the collapse-time distribution will be generalized to the two-point joint distribution, allowing us to see how knowledge of the collapse time at one radius affects the distribution of collapse times for other radii. Finally a mapping will be introduced converting the time elapsed along an observer's world line to the increase in radius out to which they can see. The result of all of this will be a statement about what it is our descendants might perceive the universe's properties to be, given what it is we currently know about those properties.

The principle collapse model to be used in this document is based on the Tolman-Lemaître line element. This description of the geometry of space-time provides a fully relativistic treatment of irrotational, spherically symmetric, cosmological perturbations. This is not, however, the line element traditionally used to describe linear cosmological perturbations of the sort generated during the inflation epoch. In this document, the so-called linear Newtonian line element will be used for this and it will be necessary to construct matching conditions to "connect" the final state of the Newtonian line element to the initial state of the Tolman-Lemaître line element for the purpose of determining collapse times. The assumption of sphericity and the absence of rotation in large cosmological perturbations limits the validity of the collapse model, and the practical implications of this are addressed in Chapter 5.

A second collapse model will briefly be used to study the effect of a cosmological constant on the results. This model is appropriate for describing a spherical ball of dust embedded in an inflating background. Observational evidence now suggests that our universe is currently undergoing a period of inflation. This raises the question of how this inflation influences the distribution of gravitationally bound objects.

Throughout this document, Planck units will be used. In this system of units, $\hbar = G = c = 1$. The sign convention used is the Landau-Lifshitz space-like convention, or $(+++)$ in the notation of Misner-Thorne-Wheeler, [65]. See Appendix A for more details on the sign conventions and units used in this document.

Chapter 2

Cosmological Inflation

2.1 Friedmann-Robertson-Walker Model

Let's begin by reviewing the essential elements of cosmological inflation. Apart from providing an introduction to the subject matter, this will provide definitions for the quantities used throughout the rest of this document and also illustrate the relationships between them. The first step in reviewing inflationary cosmology is to review the Friedmann-Robertson-Walker model of the cosmos. We begin, of course, with Einstein's field equation,

$$G_{\mu\nu} = 8\pi T_{\mu\nu}, \tag{2.1.1}$$

which is a second-order non-linear differential equation defining the relationship between the metric tensor for the geometry of space-time and the material content of space-time, as described by the stress-energy tensor $T_{\mu\nu}$. Our task is to find a solution to this equation which matches the observed gross features of our universe. Solutions to Einstein's field equation are terribly difficult to find so we will use some guessing to short-cut the process. We will do two things: assume a form for the stress-energy tensor *and* assume a form for the metric. In assuming forms for both we are, strictly speaking, over specifying the solution and we run the risk of making inconsistent assumptions. The hope is that there will be enough free parameters left over that the two assumptions can be brought into agreement with one another. It will, of course, turn out that the assumptions are consistent, which is why they are presented here. The first assumption, then, is in regards to the form of the stress-energy tensor, $T_{\mu\nu}$. We will assume that whatever it is that's going to fill our universe, it can be described as a perfect fluid. Perfect fluids are characterized, at all points, by a four-velocity, u_μ, a mass-energy density, ρ, and a pressure, p, and have a stress-energy tensor given by[1]

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu + p g_{\mu\nu} \tag{2.1.2}$$

where $g_{\mu\nu}$ is the metric tensor for the geometry of space-time. At this point, the four velocity can be left as a free parameter of the fluid but it will be found that the Einstein tensor we obtain from our (to be) assumed metric is diagonal. In order to maintain equality, all off-diagonal components of the stress-energy tensor must also be zero, which forces the fluid to be co-moving — at rest with respect to the spatial co-ordinate grid. For simplicity we will take this to be the case from the start. This restriction sets to zero all but the time component of the fluid's four velocity.

---

[1] See [65, page 711].

We next assume a form for the metric. Since on the largest scales our universe appears to be spatially isotropic and homogeneous, we would like to use a metric with these properties. It turns out that imposing these conditions significantly narrows down the possibilities to just three metrics with each having a time-dependant scale factor as their only undetermined parameter. The three metrics represent the three possibilities for the curvature of the spatial slices — open, flat or closed — and can be put into a common form by introducing a parameter, k, which is used to enumerate them. A detailed derivation of the metric is given in [65, chapter 27]. The result is the Robertson-Walker metric,

$$\mathrm{d}s^2 = -\,\mathrm{d}t^2 + a^2(t)\left[\frac{\mathrm{d}r^2}{1-kr^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right], \tag{2.1.3}$$

where $a(t)$ is the arbitrary scale factor and $k = -1, 0, +1$ for negative, zero, and positive spatial curvatures respectively (open, flat, and closed universes). Note that at this stage we have yet to impose Einstein's field equation — this metric was obtained through purely geometric considerations. It will sometimes be useful to work with a time-like co-ordinate other than t. By factoring out $a^2(t)$ we obtain the metric

$$\mathrm{d}s^2 = a^2(\eta)\left[-\,\mathrm{d}\eta^2 + \frac{\mathrm{d}r^2}{1-kr^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right] \tag{2.1.4}$$

where η, called the conformal time or arc-parameter [65, Section 27.9], is defined by $\mathrm{d}\eta^2 = \frac{1}{a^2(t)}\,\mathrm{d}t^2$ or

$$\eta = \int_0^t \frac{1}{a(t)}\,\mathrm{d}t \tag{2.1.5}$$

and in the form of the metric shown above, $a(t)$ is indicated as being parameterized in terms of η.

If we now substitute the metric in (2.1.3) and stress-energy tensor in (2.1.2) into Einstein's equation, we will obtain equations of motion for the scale factor, $a(t)$. This constitutes the Friedmann-Robertson-Walker solution. While the calculations are straightforward, they are quite lengthy when done by hand. The result is[2]

$$2\left(\frac{\ddot{a}}{a}\right) + \left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} + 8\pi p = 0 \tag{2.1.6a}$$

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} = \frac{8\pi}{3}\rho \tag{2.1.6b}$$

---

[2] These can be found in [65, equations 27.39].

where a dot indicates differentiation with respect to t. These are sometimes called the Friedmann equations and sometimes (2.1.6b) alone is called the Friedmann equation. We have two equations in three unknowns: $a(t)$, $\rho(t)$, and $p(t)$. A solution can be arrived at when a third equation is supplied: the equation of state for the perfect fluid, relating ρ to p. In any case, one can obtain (see Appendix C.3) an energy conservation law from equations (2.1.6a) and (2.1.6b), namely[3]

$$\dot{\rho} + 3(\rho + p)\frac{\dot{a}}{a} = 0. \tag{2.1.7}$$

Typical equations of state relating ρ and p are of the form

$$p = w\rho. \tag{2.1.8}$$

Some examples: for non-relativistic cold matter, or "dust," $w = 0$; while for an ultra-relativistic noninteracting gas $w = \frac{1}{3}$.[4] Substituting $p = w\rho$ into (2.1.7), it is possible to show that (see Appendix C.4)

$$\rho \propto a^{-3(1+w)} \tag{2.1.9}$$

which, upon substitution into (2.1.6b), gives

$$\dot{a}^2 + k \propto \frac{8\pi}{3} a^{-(1+3w)}. \tag{2.1.10}$$

As long as $w \geq -\frac{1}{3}$, the k term can be made negligibly small next to the a term for sufficiently small a so dropping k from (2.1.10) and taking the positive square root (for an expanding universe) gives

$$\frac{\mathrm{d}a}{\mathrm{d}t} \propto a^{-\frac{1}{2}(1+3w)} \tag{2.1.11}$$

or

$$a^{\frac{1}{2}(1+3w)}\,\mathrm{d}a \propto \mathrm{d}t. \tag{2.1.12}$$

Integrating gives

$$a \propto (t - t_0)^{\frac{2}{3(1+w)}} \tag{2.1.13}$$

where t_0 is an arbitrary constant fixing the origin of the time co-ordinate. Without loss of generality, t_0 can be taken to be zero. Throughout this calculation it has been assumed that w is a constant. In reality, one expects w to not be a constant. For example, at early times the universe was probably filled with a hot ultra-relativistic gas which then cools as the universe expands and eventually looks more like cold non-relativistic dust. If w is not constant in time then the behaviour of $a(t)$ is not that shown in (2.1.13).

The factor $\dot{a}/a$ that appears in equations (2.1.6a) and (2.1.6b) is the Hubble constant, H. This describes the rate of expansion of the universe, being in units of rate of change of length per unit length. For a flat, $k = 0$, universe there is a simple relationship between

---

[3] Compare to [60, equation 1.3.8].

[4] See [60, page 14].

H and ρ given by (2.1.6b), namely

$$\rho_c \equiv \rho = \frac{3}{8\pi} H^2, \tag{2.1.14}$$

which defines ρ_c or "critical density": the density one would observe in a flat universe. It proves convenient to introduce the dimensionless parameter Ω, being the ratio of the universe's actual energy density to the critical density,

$$\Omega = \frac{\rho}{\rho_c}. \tag{2.1.15}$$

By writing ρ as

$$\rho = \Omega \rho_c = \frac{3}{8\pi} H^2 \Omega, \tag{2.1.16}$$

(2.1.6b) becomes

$$\left(\frac{\dot{a}}{a}\right)^2 (1 - \Omega) + \frac{k}{a^2} = 0 \tag{2.1.17}$$

with

$$k = \operatorname{sign}(\Omega - 1) = \frac{\Omega - 1}{|\Omega - 1|}. \tag{2.1.18}$$

Today, the evidence is that our universe is very close to being critical, $\Omega \approx 1$.[5] Using (2.1.18) in (2.1.17) allows us to do a quick investigation of the behaviour of Ω as a function of time.

$$\left(\frac{\dot{a}}{a}\right)^2 (1 - \Omega) - \frac{1}{a^2} \frac{1 - \Omega}{|\Omega - 1|} = 0 \tag{2.1.19}$$

or, for $\Omega \neq 1$,

$$|\Omega - 1| = \dot{a}^{-2}. \tag{2.1.20}$$

(2.1.13) tells us that for $|\Omega - 1| \ll 1$, $a(t) \propto t^{\frac{2}{3(1+w)}}$, so $\dot{a}^{-2} \propto t^{\frac{2}{3}\frac{1+3w}{1+w}}$, and so (2.1.20) becomes

$$|\Omega - 1| \propto t^{\frac{2}{3}\frac{1+3w}{1+w}}. \tag{2.1.21}$$

From (2.1.13) we also know how the age of an $|\Omega - 1| \ll 1$ universe with a constant w is related to the Hubble constant, namely

$$t = \frac{2}{3(1+w)H}.$$

Considering a universe composed entirely of cold non-relativistic gases, that is $w = 0$, and using the data in Table B.5 for the Hubble constant we get[6]

$$t = \frac{2}{3H} = 14\ \text{Ga}. \tag{2.1.22}$$

---

[5] See Appendix B for current numerical data.

[6] Compare to [60, page 14].

This is in fairly good agreement with the observed lower bounds for the age of the universe considering the coarseness with which it was obtained. Combining this age with (2.1.21) tells us that $\Omega - 1$ has increased by a factor of about 3×10^{40} since $t = 1\ t_{\mathrm{Pl}}$. This result was obtained assuming a cold non-relativistic matter-dominated universe right back to the big bang which, of course, is not realistic. A more thorough analysis considering the effects of an early period of radiation domination suggests an increase on the order of 10^{55}.[7] The significance of this is seen when one recalls that the observed value of Ω is very close to 1 today; in order for Ω to be as close to 1 as it is today, the universe had to depart from being exactly critical by no more than 1 part in 10^{55} at the time of the big bang. Had the difference been much more either way, the universe would have either re-collapsed long ago or have diluted so rapidly that the matter density would not have been large enough, for long enough, for life as we know it to have developed. This remarkable feat of fine-tuning suggests that perhaps there was a physical reason for Ω to have been so close to 1 at the outset.

Notice from (2.1.20) that if $\ddot{a} > 0$, that is if $\dot{a}$ were to increase with time rather than decrease, as above, then Ω would be driven *toward* 1 rather than away from it. This can be achieved if the equation of state for the dominant component of the cosmological fluid differs from what we've used above. Any period of the universe's life in which $\ddot{a} > 0$ is now referred to as a period of "cosmological inflation."

Since the Hubble constant determines both the age of the universe and its expansion rate, it also determines the size of the observable universe by putting a limit on the volume of space that can be causally connected to an observer. The boundary of this volume is known as the particle horizon, and its distance from us is computed as follows.[8] A photon travelling along a radial null geodesic in the Robertson-Walker geometry follows a path described by

$$\mathrm{d}t^2 = a^2(t)\frac{\mathrm{d}r^2}{1 - kr^2} \tag{2.1.23}$$

which is obtained by setting $\mathrm{d}\theta$, $\mathrm{d}\phi$, and $\mathrm{d}s$ to 0 in (2.1.3). For a $k = 0$ Friedmann-Robertson-Walker universe in which $a(t) = a_0 t^n$ with $n < 1$,

$$r(t) = \int_0^t \frac{\mathrm{d}t}{a(t)} = \frac{1}{a_0}\frac{t^{1-n}}{1-n}. \tag{2.1.24}$$

This co-ordinate distance corresponds to a proper distance of

$$R_{\mathrm{p}} = a(t)r(t) = (1-n)^{-1}t. \tag{2.1.25}$$

Note that the computation of (2.1.24) relies on knowing the time dependence of $a(t)$ right back to $t = 0$ which this simple power-law approximation does not model correctly. In fact, for $n \sim 1$, the result relies very strongly on the time dependence while for $n \geq 1$, R_{p} becomes infinite.

We are living in a very flat, very old, yet relatively dense universe. In order for an

---

[7] See [60, Section 1.5].

[8] This can be found in [72, Section 2.6].

FRW universe to appear as dense as ours does after being around for as long as ours has it is necessary for the initial density to be equal to the critical density to better than 1 part in 10^{55}. Therefore, although it is possible to fit the big bang cosmological model to observation, it seems to only be possible to do so through extraordinarily precise fine-tuning of the initial conditions [48, page 266].[9] The problem of fine-tuning the initial conditions can be eliminated to some extent by considering the effect of filling the early universe with some material having a wildly different equation of state from that ones considered above. The result is an accelerating scale factor — a period of cosmological inflation.

2.2 Behaviour Of Scalar Fields

The early work on inflationary models of the cosmos was carried out, among others, by Starobinsky [85], Guth [30], and Linde [58, 59]. The advantage of these models is that they circumvent many of the problems of the standard hot big bang cosmology. Specifically, in inflationary models the choice of initial conditions is almost completely arbitrary as almost all choices rapidly evolve toward the conditions that we see in our universe today.

A modern description of inflation begins with a discussion of scalar fields and their behaviour under the conditions of the very early universe. Here I'll consider a massive, real, scalar field with the simplest coupling to gravity which is described by the Lagrange density[10]

$$\mathcal{L} = L\sqrt{-g} = -\frac{1}{2}\left[g^{\mu\nu}\phi_{;\mu}\phi_{;\nu} + m^2\phi^2\right]\sqrt{-g} \tag{2.2.1}$$

where ϕ is the field variable, and m is the mass of the field quanta. Using the Euler-Lagrange equation, one can obtain the equation of motion for this field. This is done in Appendix C.5 and the result is

$$\Box\phi = m^2\phi \tag{2.2.2}$$

where $\Box$ is the covariant d'Alembertian, $\Box\phi = \phi^{;\mu}{}_{;\mu}$. We will also need the stress-energy tensor for this field. This is given by[11]

$$T_{\mu\nu} = \phi_{;\mu}\phi_{;\nu} - \frac{1}{2}\phi^{;\alpha}\phi_{;\alpha}g_{\mu\nu} - \frac{1}{2}m^2\phi^2 g_{\mu\nu} \tag{2.2.3}$$

So far we have been considering the behaviour of a universe filled with a perfect fluid and the results obtained were quite simple and straight forward. It would be nice if these results could be used to study the behaviour of a universe filled instead with the scalar field described by (2.2.1). The trick, then, is to find a way by which the scalar field's stress-energy tensor in (2.2.3) can be made to look like that of a perfect fluid in (2.1.2).

---

[9] In the context of an initial cosmological singularity, the "initial conditions" are the conditions that exist at the time when a classical description of space-time first becomes reasonable; typically a small number of Planck-times after the singularity.

[10] The phenomenon of inflation does not rely on the field being a scalar. It can be achieved using other models. See, for example, [47] and [60, Chapter 9].

[11] See [12, equation (3.190)] with $\xi = 0$. It should be noted that although their final expression for the tensor is correct, the variational identities Birrell and Davies claim to have used to obtain it are not all correct.

If we assume that the scalar field is sufficiently uniform that spatial derivatives of ϕ can be neglected from (2.2.3), we are left with

$$\begin{aligned} T_{\mu\nu} &\approx \dot{\phi}^2 \delta^t{}_\mu \delta^t{}_\nu + \left(\frac{1}{2}\dot{\phi}^2 - \frac{1}{2}m^2\phi^2\right) g_{\mu\nu} \\ &= \dot{\phi}^2 \delta^t{}_\mu \delta^t{}_\nu + \left(\frac{1}{2}\dot{\phi}^2 - V(\phi)\right) g_{\mu\nu}. \end{aligned} \tag{2.2.4}$$

Comparing this to (2.1.2) one sees that this is the stress-energy tensor for a perfect fluid with

$$\rho = \frac{1}{2}\dot{\phi}^2 + V(\phi), \tag{2.2.5a}$$

$$p = \frac{1}{2}\dot{\phi}^2 - V(\phi). \tag{2.2.5b}$$

For a sufficiently slowly varying field, this is equivalent to the equation of state

$$p = -\rho = -V(\phi), \tag{2.2.6}$$

or $w = -1$. What sort of perfect fluid is this? Substituting this equation of state into the energy conservation equation in (2.1.7) gives $\dot{\rho} = 0$ so the energy density of such a fluid, when filling an FRW universe, remains constant regardless of the behaviour of the scale factor. In fact, from the equation of state in (2.2.6), one can see that such a fluid should have this property since any work done against the pressure to change the fluid's volume exactly equals the change in total energy required to maintain the fluid's energy density at a constant level. Another property of perfect fluids is that they have no viscosity. This stress-energy tensor, therefore, describes a substance that has no viscosity and an energy density that's independent of its volume — properties of a vacuum. For this reason, this kind of fluid is known as a "pseudo-vacuum."[12]

2.3 Application To Cosmology

Gross Features

Let us now imagine filling a Friedmann-Robertson-Walker universe with a sufficiently uniform and slowly-varying scalar field — a pseudo-vacuum. The behaviour of the universe is given by (2.1.6b) with ρ set to a constant. This equation can be solved explicitly for all three choices of spatial curvature: open, closed and flat. One can verify the following results for $k \neq 0$ by direct substitution. If $k = 1$ then from (2.1.6b) we have

---

[12] For a discussion of inflation starting from the assumption of a non-vanishing vacuum energy density, rather than from the assumption of the presence of this or that scalar field, see [29].

$$\left(\frac{8\pi}{3}\rho\right) a^2 - \dot{a}^2 = 1 \tag{2.3.1}$$

and this is solved by $a(t) = H^{*-1} \cosh H^* t$ with $H^* = \sqrt{\frac{8\pi}{3}\rho}$. Likewise, for $k = -1$ one finds that $a(t) = H^{*-1} \sinh H^* t$ is the solution. The use of the symbol H^* in these solutions is suggestive of the Hubble constant but for the moment consider it to be merely a constant whose value is as given above. More will be said about this below. Finally, if $k = 0$ then (2.1.6b) becomes

$$\dot{a} = \pm\sqrt{\frac{8\pi}{3}\rho}\, a. \tag{2.3.2}$$

For an expanding universe, this is solved by $a(t) = \beta e^{H^* t}$ with the same expression for H^* as above and with β an arbitrary constant having the same dimensions as H^{*-1} i.e. length.[13] Choosing $\beta = H^{*-1}$ for consistency with the other two cases, then altogether we have

$$a(t) = H^{*-1} \begin{cases} \cosh H^* t & \text{if } k = +1, \\ e^{H^* t} & \text{if } k = 0, \\ \sinh H^* t & \text{if } k = -1. \end{cases} \tag{2.3.3}$$

In all three cases, the solution of the equation of motion for $a(t)$ is characterized by exponential growth.[14] These results are different from (2.1.13) because that result only holds for $w > -\frac{1}{3}$ while here $w = -1$.

Note that while the solutions given in (2.3.3) are exact, the H^* appearing in the solutions for the $k = \pm 1$ cases is not exactly equal to the Hubble constant; that is for $k = \pm 1$,

$$H = \frac{\dot{a}}{a} = H^* \left(\tanh H^* t\right)^{\pm 1} \neq H^*. \tag{2.3.4}$$

The choice of symbol is justified, however, because for $H^* t > 1$, H for the $k = \pm 1$ cases can be approximated by

---

[13] One has to pay attention here. Clearly (2.3.2) cannot determine the dimensionality of β. Instead, one must recall that (2.3.2) was originally $\text{RHS}^2 - \text{LHS}^2 = k = 0$ where k is dimensionless. It is from this and the form of the solution for $a(t)$ that we obtain the restriction on β. Despite this, one will sometimes see other choices being made in the literature which result in minor changes in notation. By giving $a(t)$ the dimensions of H^{-1} we are choosing the spatial co-ordinates to be dimensionless. This means, for example, that the conformal time, η, in (2.1.4) must also be dimensionless. In [60, Section 7.3] Linde instead gives η the dimension of length by saying the conformal time for $k = 0$ de Sitter space is $\eta = -H^{-1}e^{-Ht}$. This means that he is implicitly choosing $a(t)$ to be dimensionless. However, in [60, Section 1.6], Linde had *already* given $a(t)$ dimensions of length when he said $a(t) = H^{-1}e^{Ht}$. This inconsistency should cause him to pick up a factor of $H^{3/2}$ in [60, equation (7.3.6)] which he does not so he, himself, is being careful enough to not get tripped-up by this inconsistency. By missing this subtlety in choice of dimensionality one can pick up factors of H when mixing expressions from different authors or from different points in the same author's work.

[14] See also [60, pages 34–35].

$$H = \frac{\dot{a}}{a} = H^* \left(\tanh H^* t\right)^{\pm 1} \approx H^* \left(1 \mp 2\mathrm{e}^{-2H^* t}\right) \tag{2.3.5}$$

respectively so for $H^* t \gtrsim 1$, $H = \frac{\dot{a}}{a}$ quickly asymptotes to H^*. More justification comes from examining the rate of change of the Hubble parameter. Starting with the $k = \pm 1$ cases in (2.3.3),

$$\frac{\ddot{a}}{a} = H^{*2} \tag{2.3.6}$$

or

$$\left(\frac{\dot{a}}{a}\right)^{\cdot} = H^{*2} - \left(\frac{\dot{a}}{a}\right)^2 = \left(\frac{\dot{a}}{a}\right)^2 \left(\tanh^{\mp 2} H^* t - 1\right), \tag{2.3.7}$$

where (2.3.5) has been used, so

$$\frac{\dot{H}}{H^2} = \left(\frac{\dot{a}}{a}\right)^{\cdot} \Big/ \left(\frac{\dot{a}}{a}\right)^2 \approx \pm 4\mathrm{e}^{-2H^* t},$$

or

$$\frac{\dot{H}}{H^2} \ll 1 \tag{2.3.8}$$

for $H^* t \gg 1$.[15] $\dot{H}/H^2$ is the fractional of change of the Hubble constant per Hubble time — the fractional change in the e-fold time per e-fold time — so this quantity being much less than one implies the expansion is very close to exponential. Likewise, from (2.3.7), the second time derivative of the Hubble constant is

$$\ddot{H} = \left(\frac{\dot{a}}{a}\right)^{\cdot\cdot} = -2\left(\frac{\dot{a}}{a}\right)\left(\frac{\dot{a}}{a}\right)^{\cdot} = -2\left(\frac{\dot{a}}{a}\right)^3 \left(\coth^2 H^* t - 1\right),$$

or

$$\frac{\ddot{H}}{H^3} \ll 1 \tag{2.3.9}$$

for $H^* t \gg 1$. The two relationships in (2.3.8) and (2.3.9) will be used repeatedly to approximate expressions during the inflation epoch. Finally, since H approaches H^* very quickly, the condition $H^* t \gg 1$ can be replaced with $Ht \gg 1$ for any of the results above.

In summary, if the universe is filled with a sufficiently uniform scalar field, then very quickly (within a few multiples of H^{*-1}) it does not matter if the universe is open, closed, or critical; its behaviour is approximately the same and that behaviour is the exponential increase of its size with time — the universe is approximately de Sitter. The only difference between the three scenarios is a small correction to the exponential growth.

---

[15] Compare this expression and the result that follows to the statements made in [60, page 35] but note that Linde confusingly switches between H the constant quantity appearing in (2.3.3) and H the Hubble "constant." These are different quantities but sufficiently similar to warrant approximating them as being equal in certain contexts. Of course, this is the point he is trying to make.

Details

Now that we have a rough picture of the behaviour of this variant of the standard FRW cosmos, let's perform a somewhat more quantitative analysis. Using $\rho = V(\phi) + \frac{1}{2}\dot{\phi}^2$ from (2.2.5a) in (2.1.6b), one of the Friedman equations, gives

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} \equiv H^2 + \frac{k}{a^2} = \frac{8\pi}{3}\left(V(\phi) + \frac{1}{2}\dot{\phi}^2\right). \tag{2.3.10}$$

The field, ϕ, must also satisfy its equation of motion, (2.2.2), which can be written as

$$\ddot{\phi} + 3\frac{\dot{a}}{a}\dot{\phi} = -\frac{\mathrm{d}V}{\mathrm{d}\phi}. \tag{2.3.11}$$

See Appendix C.5 for the derivation. Clearly, if $\dot{a} > 0$ (i.e. the universe is expanding) then eventually a will become large and the k/a^2 term can be dropped from (2.3.10) giving

$$H^2 = \frac{8\pi}{3}\left(V(\phi) + \frac{1}{2}\dot{\phi}^2\right) \tag{2.3.12}$$

or, in the "slow-roll" approximation, $\dot{\phi}^2 \ll V(\phi)$,

$$H^2(\phi) \approx \frac{8\pi}{3}V(\phi). \tag{2.3.13}$$

Here, H has been written as $H(\phi)$ indicating that this relationship is intended to be used as an approximate expression for the Hubble constant during the inflation epoch. This approximation will be used repeatedly. From this point on in the document, it will be assumed that our universe (or the background, once perturbations are considered) is accurately modelled by a flat FRW cosmos so that where the choice would matter in some expression, we will take the $k = 0$ one. Applying the slow-roll approximation, $\ddot{\phi} \ll dV/d\phi$, to (2.3.11) and substituting (2.3.13) into the result gives[16]

$$\frac{1}{2}\dot{\phi}^2 = \frac{1}{48\pi V}\left(\frac{\mathrm{d}V}{\mathrm{d}\phi}\right)^2 \tag{2.3.14}$$

which, for $V \propto \phi^n$, is

$$\frac{1}{2}\dot{\phi}^2 = \frac{n^2}{48\pi\phi^2}V(\phi). \tag{2.3.15}$$

This expression shows for large ϕ, i.e.

$$\phi \gg \frac{n}{4\sqrt{3\pi}}, \tag{2.3.16}$$

---

[16] Note the error in [60, equation (1.7.17)] and the fact that they are using units in which $m_{\mathrm{Pl}} \neq 1$.

that[17]

$$\dot{\phi} \ll V(\phi). \tag{2.3.17}$$

Apart from being circularly-consistent with our specification that $T_{\mu\nu}$ should be dominated by $V(\phi)$ (assumption of homogeneity plus "slow-roll" approximation), this result gives us some quantitative information regarding the domain of ϕ for which the slow-roll approximation can be expected to hold.

At this point, we can also make a statement about how much the universe grows during the inflation epoch while ϕ decays. Starting by rearranging (2.3.15) into the form

$$V(\phi) = \frac{24\pi}{n^2}\phi^2\dot{\phi}^2 \tag{2.3.18}$$

and substituting this into (2.3.13) gives

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}\frac{24\pi}{n^2}\phi^2\dot{\phi}^2$$

which can be integrated to give

$$a(t) = a_0 \exp\left[\frac{4\pi}{n}\left(\phi_0^2 - \phi^2(t)\right)\right] \tag{2.3.19}$$

where the negative choice of square-root has been made in order to get $a(t)$ to grow as t increases and $\phi(t)$ decays.[18] We see, then, that over the course of the inflation epoch the Robertson-Walker scale factor increases by as many e-folds as the change in the square of the field amplitude. One consequence of this is that regions with higher initial ϕ_0 are responsible for a strongly disproportionate fraction of the final volume of the universe. An observer chosen at random, then, can be fairly certain that their visible patch of the universe started out with a very large initial value for the inflaton field.

One final remark that can be made here is with regard to the value of ϕ at the end of inflation. In order for the expansion of the universe to be quasi-exponential, it is necessary for $V(\phi)$ to dominate over spatial and time derivatives of ϕ in $T_{\mu\nu}$. We need the range of values for ϕ for which the above is true and for this we will simply take (2.3.16) to be a bound specification. In other words, inflation will be said to have ended when

$$\phi = \frac{n}{4\sqrt{3\pi}}. \tag{2.3.20}$$

An illustration of the behaviour of $\phi(t)$ is given in Figure 2.1 along with the corresponding behaviour of $H(t)$. As can be seen, in both models the Hubble constant experiences a decline throughout the inflation epoch. In the $m^2\phi^2$ model, H decays linearly during the inflation epoch while in the $\lambda\phi^4$ model the decay is exponential. The behaviour of both models is plotted out to within the start of the "ring-down" phase of the system. The accuracy of (2.3.20) in determining the end of the inflation epoch is clearly visible

---

[17] Compare to [60, page 45].

[18] Compare to [60, equation (1.7.25)].

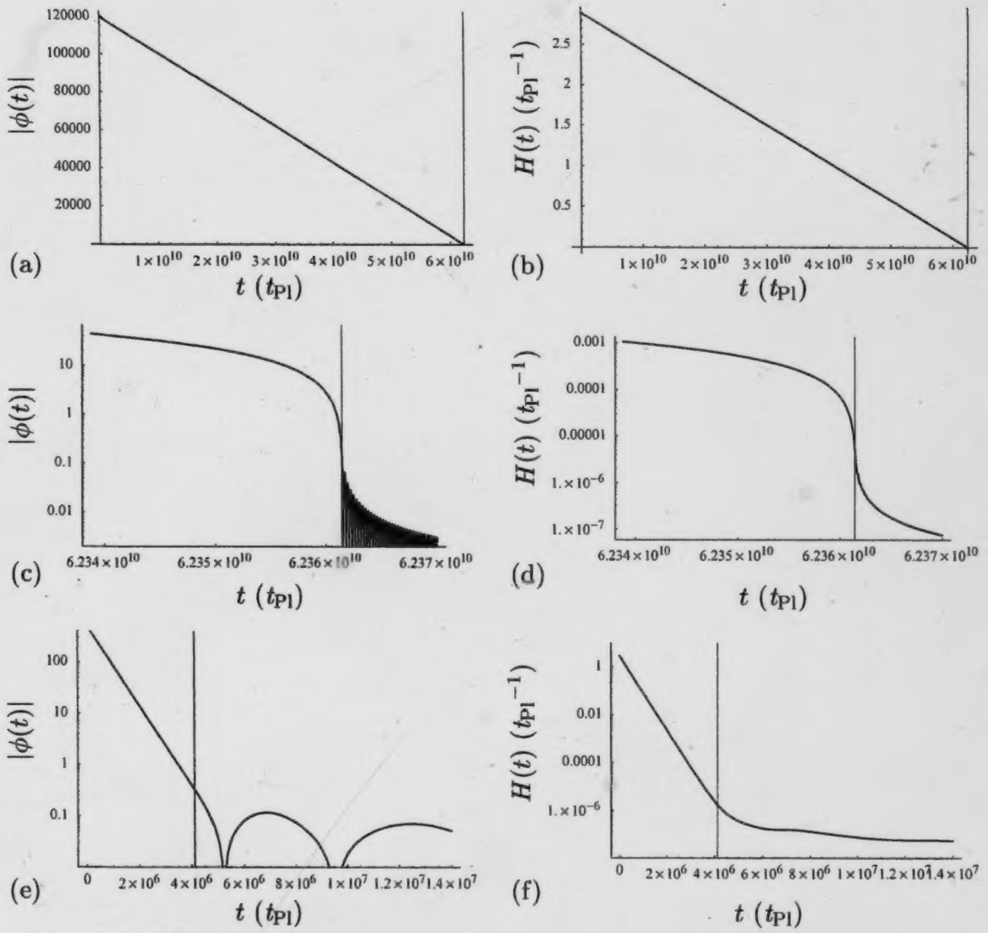


Figure 2.1: Numerical integration of (2.3.11) and (2.3.12) showing the typical time dependence of ϕ and H during and immediately following the inflation epoch. In all cases, the system was started with the initial conditions $V(\phi) = 1$, $\ddot{\phi} = 0$ and $\dot{\phi} < 0$. The horizontal axis in each plot is time in units of t_{Pl} and is marked off at the specific time at which (2.3.20) is true. (a) $\phi(t)$ for the $\frac{1}{2}m^2\phi^2$ potential with $m = 1.18 \times 10^{-5}$. (b) $H(t)$ for the same model. (c) An expanded view of the curve in (a) at the end of the inflation epoch. (d) An expanded view of the curve in (b) at the end of the inflation epoch. (e) $\phi(t)$ for the $\frac{1}{4}\lambda\phi^4$ potential with $\lambda = 6 \times 10^{-11}$. (f) $H(t)$ for the same model.

in this plots.

To get an idea of the numbers involved, consider that in a $\frac{1}{2}m^2\phi^2$ theory, the need for the homogeneity of the universe, $\frac{\delta\rho}{\rho}$, to be about 10^{-5} leads to a value of $m \approx 10^{-5}$ (this will be shown below but take it as given for now). This means that to have $V(\phi_0) \approx 1$ requires $\phi_0 \approx 10^5$. Inflation ends when $\phi \approx 1$, so over the course of inflation the scale factor increases by about 10^{10} e-folds. If the original patch is just 1 Planck length across, the final volume will be $e^{10^{10}} l_{\mathrm{Pl}}$ across.[19] The Hubble diameter of our universe is only[20] $e^{141} l_{\mathrm{Pl}}$ so this hypothetical patch will have inflated to a size that, in a statistical sense, is practically infinite — the patch is so large that statistics taken over its final volume, of properties measured on scales matching that of our observable universe, will differ negligibly from those taken over a truly infinite domain. In other words, the final patch is large enough to allow it to be taken as "ergodic."[21]

To summarize the results so far: if, by chance, some region of the initially random universe looks sufficiently de Sitter-like,[22] then a scalar field whose Lagrangian is given by (2.2.1) will, within that region, behave like a pseudo-vacuum and the region will experience exponential growth. This exponential inflation of space-time acts as a friction term in (2.3.11), the equation of motion for ϕ, preventing ϕ from "rolling" down its potential to the true vacuum state. This allows the expansion to proceed for an extended period of time.

The expansion does, however, slow over time and this reduces the friction term. As the friction decreases, the rate at which the field approaches the true vacuum increases. Eventually the field becomes close enough to the true vacuum that the expansion (and friction term) becomes negligible. The field then oscillates at the minimum of its potential. To continue the analysis farther, it is necessary to consider the coupling that should exist between the inflaton field and other particle fields. Through these couplings, the inflaton oscillations will excite modes in the other fields. The energy originally stored in the pseudo-vacuum is thus transformed into a bath of particles and the universe becomes filled with an extremely hot gas.[23] At that point the universe looks very much like the early stages of the standard hot big bang model except that it has been stretched out to an enormous size, typically *much* larger than the currently visible volume. In so doing, it has become extremely flat and homogeneous and any relics from the original creation have become exponentially diluted.

---

[19] The numbers are not quite as spectacular for the $\frac{1}{4}\lambda\phi^4$ model where the universe ends up being a modest $e^{10^5} l_{\mathrm{Pl}}$ across.

[20] See Table B.4.

[21] Ergodicity is discussed in Chapter 3.

[22] Whether or not any such regions exist or are likely to exist depends on the initial quantum state of the universe. For the inflation scenario to have a bearing on the early universe, one must assume that the quantum state is such that the probability of such a region existing is non-negligible. The issue of whether or not inflation can get started in an inhomogeneous universe is sometimes referred to as the "problem of isotropization". For analytic and numerical investigations of this problem, see for example [17], [21], [26], and [27].

[23] The process by which energy from the inflaton field is dumped into other matter fields is called "reheating" and is a subject of research unto itself. For more information, see the work by Kofman, Linde and Starobinsky in [44], [45], and [46].

Although very successful at explaining the initial conditions of the hot big bang model, inflationary models come with their own problems. The most significant of these is that in order to get the models to work right, one must often fine-tune the parameters of the Lagrangian very precisely. This leaves the models open to the possibility that the only parameter choices possible for inflation will be inconsistent with the parameter choices required in order to bring the theory into agreement with particle physics experiments. In fact, the particular choices of parameters required are often unrealistic [60, pages 38–39].

Chapter 3

The Description of Cosmological Perturbations

Our discussion of cosmic structure will now turn to perturbations. Before dealing directly with perturbations to the Friedmann-Robertson-Walker cosmos, however, we must develop some descriptive mathematics. The problem, here, is that there does not exist in the cosmology literature a single uniform notation and nomenclature for describing cosmological perturbations. One appears to be forming but at present the inhomogeneity is quite severe and can lead to a great deal of confusion and error when comparing the works of different authors. It is for this reason important that a detailed and consistent framework be developed for use within this document and it is this task that we now turn to.

This chapter is, roughly, split into four parts. At first, a generic introduction to the theory of Gaussian random fields will be presented. This is followed by a brief explanation of the mapping between the parameters used to describe random fields and those found in the cosmology literature. Next, an analysis is presented of the effects of smoothing random fields. Finally, some additional notation is introduced specifically for cosmological matter fluctuations.

3.1 Gaussian Random Fields

The following discussion of the properties of Gaussian random fields follows information found in [7], [39], [41], [92, Chapter I Section 6], and [93].[1] Unfortunately, in this subject matter there are almost as many choices of notation as there are authors and so it is not possible to write things down in a manner that is consistent with all of them. In some cases, in fact, it appears to be impossible to come up with a choice of notation that is even consistent with all of a single author's expressions. Because of this, a great deal of what follows differs from what is found in the literature. The presentation given here is a hybrid treatment that has been tailored to the subject at hand, with a few original pieces added where required. I have attempted to construct a choice of notation that either agrees with or is as similar as possible to the majority of the texts.

---

[1] The book by Vanmarke, [93], is particularly excellent and deserves special mention.

An n-dimensional random field, $f(\vec{r})$, is a set of random variables, one for each point $\vec{r}$. The field and all of its statistical properties are entirely defined by the set of finite dimensional joint probability distribution functions,

$$P\left[f(\vec{r}_1),\ldots,f(\vec{r}_m)\right]\,\mathrm{d}f(\vec{r}_1)\cdots\mathrm{d}f(\vec{r}_m), \tag{3.1.1}$$

giving the probability that for the given $\vec{r}_i$, $i = 1,\ldots,m$, the field f simultaneously takes on values at the points $\vec{r}_i$ in the ranges $f(\vec{r}_i)$ to $f(\vec{r}_i) + \mathrm{d}f(\vec{r}_i)$. A Gaussian random field is one for which all the m-point probability distributions are Gaussian in the $f(\vec{r}_i)$.

A Gaussian m-point joint probability distribution for the random variables x_i is given by[2]

$$P(x_1,\ldots,x_m)\,\mathrm{d}x_1\cdots\mathrm{d}x_m = \frac{1}{\sqrt{2\pi}^m\sqrt{\det\mathbb{B}}}\exp\left[-\frac{1}{2}\Delta\vec{x}^{\mathrm{T}}\mathbb{B}^{-1}\Delta\vec{x}\right]\mathrm{d}x_1\cdots\mathrm{d}x_m \tag{3.1.2a}$$

where $\mathbb{B}$ is the covariance matrix of the x_i,

$$\mathbb{B}_{ij} \equiv \langle\Delta x_i\Delta x_j\rangle\,, \tag{3.1.2b}$$

and $\Delta\vec{x}$ is the column vector of residuals

$$\Delta x_i \equiv x_i - \langle x_i\rangle\,. \tag{3.1.2c}$$

The elements of the covariance matrix are also known as the second cumulants of the x_i:[3]

$$\langle\langle x_i x_j\rangle\rangle \equiv \langle\Delta x_i\Delta x_j\rangle\,. \tag{3.1.3}$$

Note that the above shows that for Gaussian random variables, any m-point joint distribution requires only the means and two-point covariance matrix for its construction.

For a random field, $f(\vec{r})$, with continuous co-ordinates the covariance matrix generalizes to the two-point covariance function,

$$B(\vec{r}_i,\vec{r}_j) = \langle\langle f(\vec{r}_i)f(\vec{r}_j)\rangle\rangle\,, \tag{3.1.4}$$

and, as above, this contains all the information required to construct any m-point joint distribution for the field.

A strictly homogeneous random field is one for which all finite-dimensional joint probability distributions, (3.1.1), are left unchanged by a co-ordinate translation.[4] This implies that the joint distribution functions depend on the co-ordinate separations, $\vec{r}_{ij} = \vec{r}_j - \vec{r}_i$, called the lag factors, alone; and that the mean, $\langle f(\vec{r})\rangle$, and mean square, $\langle f^2(\vec{r})\rangle$, are independent of $\vec{r}$. In this case, the two-point covariance function can be translated to the origin without loss of generality,

---

[2] See [93, equation (2.5.11)]. See also [92] for a different, although in this case less useful, representation.

[3] The double angle-bracket notation is not standard but is used by van Kampen in [92] due to its convenience when dealing with multiple variables and I have chosen to carry the notation through into the present document. For the definitions of cumulants, see [25, equations (2.7.4), (2.7.5)].

[4] In contrast to a homogeneous field in the "wide" sense which requires only the one- and two-point distributions to be invariant under translation.

$$B(\vec{r}_i, \vec{r}_j) = \langle\!\langle f(\vec{r}_i) f(\vec{r}_j)\rangle\!\rangle = \langle\!\langle f(0) f(\vec{r}_j - \vec{r}_i)\rangle\!\rangle = \langle\!\langle f(0) f(\vec{r})\rangle\!\rangle = B(\vec{r}), \tag{3.1.5}$$

thereby becoming a function of a single co-ordinate. If the mean, in addition to being independent of $\vec{r}$, is also 0 then the covariance function reduces to

$$B(\vec{r}) = \langle\!\langle f(0) f(\vec{r})\rangle\!\rangle = \langle f(0) f(\vec{r})\rangle \,. \tag{3.1.6}$$

In this case, this function is also sometimes denoted as $\xi(0, \vec{r}) = \xi(\vec{r})$, the (two-point) correlation function.[5]

A random field is isotropic if the joint probability distribution functions are left unchanged by rotations. For a homogeneous random field, isotropy means the distribution functions depend only on the magnitudes of the lag factors, $|\vec{r}_{ij}| = |\vec{r}_j - \vec{r}_i|$, not their direction. Finally, a random field is called ergodic if a single realization of the field contains sufficient information to completely determine all the joint probability distribution functions.

It is important to understand that when dealing with random fields one is always discussing the properties of the statistics of the field. This contrasts with classical fields where one characterizes them by the properties of their values. For this reason, the meaning of homogeneous and isotropic when used in the context of a random field is not the same as when those same words are used in the context of classical fields. In particular, a homogeneous classical scalar field is necessarily "isotropic" since a field that is the same everywhere is left unchanged by a rotation. A homogeneous random scalar field, however, can be "anisotropic" as long as the direction dependence of its statistical properties is the same everywhere.

The n-dimensional Fourier transform of the field is defined as usual and in this document its precise definition is taken to be

$$f(\vec{r}) = \frac{1}{\sqrt{2\pi}^n} \int f_{\vec{k}} \mathrm{e}^{\mathrm{i}\vec{k}\cdot\vec{r}}\, \mathrm{d}^n k, \tag{3.1.7}$$

$$f_{\vec{k}} = \frac{1}{\sqrt{2\pi}^n} \int f(\vec{r}) \mathrm{e}^{-\mathrm{i}\vec{k}\cdot\vec{r}}\, \mathrm{d}^n r. \tag{3.1.8}$$

---

[5] Recalling that the covariance function, $B(\vec{r})$ (along with the mean, $\langle f(\vec{r})\rangle$), provides all the information required to fully characterize a homogeneous Gaussian random field, one might be surprised to read in cosmology literature the same statement being made about the two-point correlation function, $\xi(\vec{r})$. See, for example, Bardeen in [7] and Mukhanov et al. in [69, Section 12.1]. These statements are not, strictly, correct. For a homogeneous field with a mean of 0 (as is most often considered in cosmology), these are identical functions. When the mean is not 0, even if the field remains homogeneous, the additional information carried by B is required. In the particular cases mentioned above, if one reads carefully one finds that Bardeen does actually restrict his statement to the case of a field with 0 mean and so while being somewhat misleading his statements are correct; but Mukhanov et al. do not make this restriction.

The Fourier transform, $f_{\vec{k}}$, is a complex random field — each $f_{\vec{k}}$ has a random modulus and argument. An important quantity is the two-point spectral correlation function. For a homogeneous random field with zero mean we can obtain this by starting from the Fourier transform in (3.1.8),

$$\begin{aligned}\left\langle f_{\vec{k}} f^*_{\vec{k}'} \right\rangle &= \frac{1}{(2\pi)^n} \left\langle \int f(\vec{r}_1) \mathrm{e}^{-\mathrm{i}\vec{k}\cdot\vec{r}_1} f(\vec{r}_2) \mathrm{e}^{+\mathrm{i}\vec{k}'\cdot\vec{r}_2} \, \mathrm{d}^n r_1 \, \mathrm{d}^n r_2 \right\rangle \\ &= \frac{1}{(2\pi)^n} \int \langle f(\vec{r}_1) f(\vec{r}_2) \rangle \, \mathrm{e}^{-\mathrm{i}\left(\vec{k}\cdot\vec{r}_1 - \vec{k}'\cdot\vec{r}_2\right)} \, \mathrm{d}^n r_1 \, \mathrm{d}^n r_2 .\end{aligned}$$

Making the change of variables

$$\begin{aligned}\vec{r} &= \vec{r}_2 + \vec{r}_1, \\ \vec{r}' &= \vec{r}_2 - \vec{r}_1, \\ \mathrm{d}^n r_1 \, \mathrm{d}^n r_2 &= \frac{1}{2^n} \, \mathrm{d}^n r \, \mathrm{d}^n r',\end{aligned}$$

gives

$$\left\langle f_{\vec{k}} f^*_{\vec{k}'} \right\rangle = \frac{1}{2^n} \frac{1}{(2\pi)^n} \int \left\langle f\left(\frac{1}{2}(\vec{r}' - \vec{r})\right) f\left(\frac{1}{2}(\vec{r}' + \vec{r})\right) \right\rangle \times \\ \mathrm{e}^{-\frac{1}{2}\mathrm{i}\left(\vec{k}\cdot(\vec{r}'-\vec{r}) - \vec{k}'\cdot(\vec{r}'+\vec{r})\right)} \, \mathrm{d}^n r \, \mathrm{d}^n r' .$$

Because the field is homogeneous, $\langle f(\vec{r}_1) f(\vec{r}_2) \rangle$ is translation invariant. This makes the expectation of the product of the f's independent of $\vec{r}'$ thus allowing us to evaluate it at whatever value of $\vec{r}'$ we please. In particular, setting $\vec{r}' = -\vec{r}$ gives

$$\begin{aligned}\left\langle f_{\vec{k}} f^*_{\vec{k}'} \right\rangle &= \frac{1}{2^n} \frac{1}{(2\pi)^n} \int \langle f(-\vec{r}) f(0) \rangle \, \mathrm{e}^{\frac{1}{2}\mathrm{i}\left(\vec{k}+\vec{k}'\right)\cdot\vec{r}} \mathrm{e}^{-\frac{1}{2}\mathrm{i}\left(\vec{k}-\vec{k}'\right)\cdot\vec{r}'} \, \mathrm{d}^n r \, \mathrm{d}^n r' \\ &= \delta^{(n)}\left(\vec{k} - \vec{k}'\right) \int B(\vec{r}) \mathrm{e}^{-\mathrm{i}\vec{k}\cdot\vec{r}} \, \mathrm{d}^n r \qquad (3.1.9)\end{aligned}$$

where, because the field's mean is 0, the mean in the integrand has been written in terms of $B(\vec{r})$. The presence of the Dirac δ-function, $\delta^{(n)}\left(\vec{k} - \vec{k}'\right)$, in this result is a statement of the statistical independence of distinct Fourier modes. Notice that this δ-function arises from the ability to separate the two integrals which is in turn a consequence of the homogeneity of the field.[6] Strictly speaking, this Dirac-δ means only that the two modes are uncorrelated, which is necessary for statistical independence but not sufficient. The field, however, is Gaussian and it can be shown that two Gaussian random variables that are uncorrelated are also independent [93, Section 2.5].

---

[6] Some authors, for example Liddle and Lyth in [55, Section 4.3], claim that the statistical independence of the Fourier modes of a field is a consequence of the field being Gaussian. This is not correct: it is homogeneity that leads to independent Fourier modes; whether the field is Gaussian or not is clearly irrelevant. It seems that in much of the cosmology literature it is believed that the term "Gaussian" implies "homogeneous" which it does not. One must pay careful attention when reading such documents to determine what, exactly, the restrictions on the field are and not assume the restrictions are as the author claims.

Let's write (3.1.9) as

$$\left\langle f_{\vec{k}} f^*_{\vec{k}'} \right\rangle = \delta^{(n)}\left(\vec{k} - \vec{k}'\right) \left\langle \left| f_{\vec{k}} \right|^2 \right\rangle \tag{3.1.10}$$

where

$$\left\langle \left| f_{\vec{k}} \right|^2 \right\rangle \equiv \int B(\vec{r}) \mathrm{e}^{-\mathrm{i}\vec{k}\cdot\vec{r}} \,\mathrm{d}^n r. \tag{3.1.11}$$

In introducing this notation for $\left\langle \left| f_{\vec{k}} \right|^2 \right\rangle$ I am moving beyond the information I have found in texts on the analysis of random fields. Another choice that would be more in keeping with the notation used throughout this document would be to set $\left\langle f_{\vec{k}} f^*_{\vec{k}'} \right\rangle = \sqrt{2\pi}^n \delta^{(n)}(\vec{k}-\vec{k}') \left\langle \left| f_{\vec{k}} \right|^2 \right\rangle$ where $\left\langle \left| f_{\vec{k}} \right|^2 \right\rangle = \frac{1}{\sqrt{2\pi}^n} \int B(\vec{r}) \mathrm{e}^{-\mathrm{i}\vec{k}\cdot\vec{r}} \,\mathrm{d}^n r$. Later this choice would lead to $B(0) = \frac{1}{(2\pi)^n} \int \left\langle f_{\vec{k}} f^*_{\vec{k}'} \right\rangle \mathrm{e}^{\mathrm{i}\vec{k}'\cdot\vec{r}} \,\mathrm{d}^n k \,\mathrm{d}^n k' = \frac{1}{\sqrt{2\pi}^n} \int \left\langle \left| f_{\vec{k}} \right|^2 \right\rangle \mathrm{e}^{\mathrm{i}\vec{k}\cdot\vec{r}} \,\mathrm{d}^n k$ instead of the expression one finds below in (3.1.14). Although this alternative choice would be more in keeping with the notation used for Fourier transforms throughout this document, it would result in a departure from the forms of the equivalent quantities one tends to find in cosmology literature. In particular, the choice of notation I have made in (3.1.10) leads to a form for (3.1.14) in which the placement of factors of 2π agrees directly with [23], [48, equation (9.14)], [55, equation (4.16)] and [72, equation (5.19)] although it disagrees in other respects with most of these.

The relationship given in (3.1.9) between the Fourier transform of the two-point spatial covariance function and the two-point spectral correlation function for homogeneous random fields with zero mean can be approached from the other direction as follows. Starting with the definition of the two-point covariance function we get

$$\begin{aligned} B(\vec{r}) &= \langle f(0) f(\vec{r}) \rangle \\ &= \frac{1}{(2\pi)^n} \left\langle \int f_{\vec{k}} f^*_{\vec{k}'} \mathrm{e}^{\mathrm{i}\vec{k}\cdot 0 + \mathrm{i}\vec{k}'\cdot\vec{r}} \,\mathrm{d}^n k \,\mathrm{d}^n k' \right\rangle \\ B(\vec{r}) &= \frac{1}{(2\pi)^n} \int \left\langle f_{\vec{k}} f^*_{\vec{k}'} \right\rangle \mathrm{e}^{\mathrm{i}\vec{k}'\cdot\vec{r}} \,\mathrm{d}^n k \,\mathrm{d}^n k'. \end{aligned} \tag{3.1.12}$$

It is easily verified that substituting (3.1.9) into this results in the identity

$$B(\vec{r}) = B(\vec{r})$$

demonstrating, at least, the mutual consistency of these two relationships. The relationships in (3.1.9) and (3.1.12) constitute a generalization of the Wiener-Khinchine relations that I have not found in texts on the subject. The more usual form of the relations can be found in [93, equations (3.2.8) and (3.2.9)]. It is important to note that in addition to generalizing the relations to off-diagonal parts of the spectral correlation function, the factors of 2π appear reversed here compared to the references I have found. As can be seen above, their present locations are due entirely to the particular definition of the Fourier transform and its inverse used here as found in (3.1.7) and (3.1.8). Further comment on this discrepancy would require more information about the definitions for Fourier transforms being used in the texts in question and this is not provided.

Setting $\vec{r}$ to 0 in (3.1.12) results in a relationship between the variance of $f(\vec{r})$ and the variance of the Fourier modes.

$$B(0) = \left\langle f^2(0) \right\rangle = \left\langle f^2(\vec{r}) \right\rangle = \frac{1}{(2\pi)^n} \int \left\langle f_{\vec{k}} f^*_{\vec{k}'} \right\rangle \mathrm{d}^n k \, \mathrm{d}^n k' \qquad (3.1.13)$$

and using (3.1.10),

$$\left\langle f^2(\vec{r}) \right\rangle = \frac{1}{(2\pi)^n} \int \left\langle \left| f_{\vec{k}} \right|^2 \right\rangle \mathrm{d}^n k. \qquad (3.1.14)$$

The form of the relation in (3.1.14) motivates the introduction of the spectral density function,[7]

$$S_{\vec{k}} = \left\langle \left| f_{\vec{k}} \right|^2 \right\rangle. \qquad (3.1.15)$$

This can be interpreted as the contribution to the variance of $f(\vec{r})$ per unit volume in $\vec{k}$-space.

Let's now consider the real and imaginary components of a field's Fourier modes along with their modulus and argument. If two independent Gaussian random variables, x_1 and x_2, both have a mean of 0 and a variance of σ^2 then the derived random variable $r = \sqrt{x_1^2 + x_2^2}$ is distributed according to the Rayleigh distribution,[8]

$$P(r) = \frac{r}{\sigma^2} \exp\left[-\frac{1}{2} \frac{r^2}{\sigma^2} \right]. \qquad (3.1.16)$$

The mean of r is

$$\langle r \rangle = \sqrt{\frac{\pi}{2}} \sigma, \qquad (3.1.17)$$

and the mean square of r is

$$\left\langle r^2 \right\rangle = 2\sigma^2. \qquad (3.1.18)$$

The mean and mean square of r are, therefore, related by

$$\langle r \rangle^2 = \frac{\pi}{2} \sigma^2 = \frac{\pi}{4} \left\langle r^2 \right\rangle. \qquad (3.1.19)$$

For a homogeneous Gaussian field, each mode of its Fourier transform, $f_{\vec{k}}$, will have real and imaginary components that are independent Gaussian random variables with means of 0 and a common variance.[9] From (3.1.18) above, therefore, their variances

---

[7] Compare to [92, equation (3.3)] and [93, equation (3.2.5)].

[8] See [93, Section 2.5].

[9] See [93, Section 3.4].

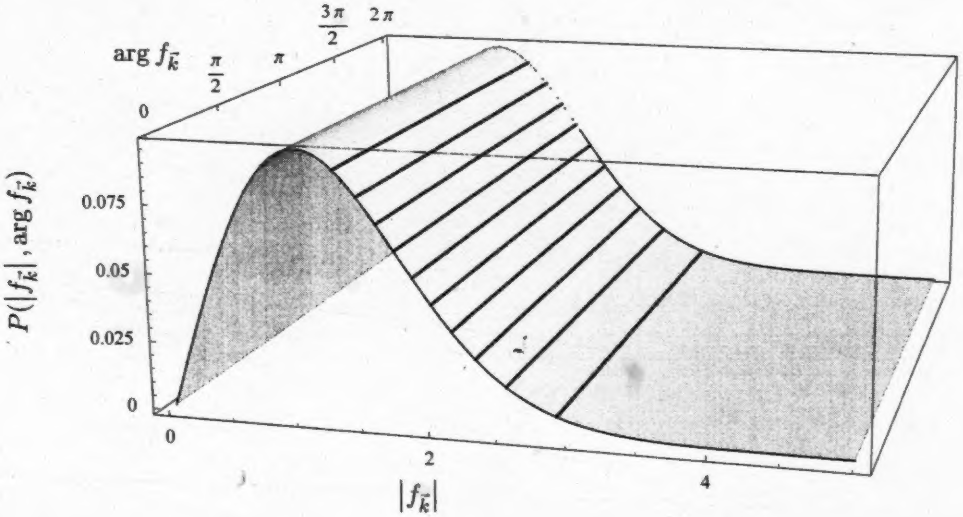


Figure 3.1: The distribution of the modulus and argument of a single mode with a spectral density of $S_{\vec{k}} = 1$ in the Fourier transform of a homogeneous Gaussian random field.

obey the relationship

$$\left\langle (\Re f_{\vec{k}})^2 \right\rangle = \left\langle (\Im f_{\vec{k}})^2 \right\rangle = \frac{1}{2} \left\langle |f_{\vec{k}}|^2 \right\rangle = \frac{1}{2} S_{\vec{k}}. \tag{3.1.20}$$

This leads to the conclusion that the phase and magnitude of each mode are random variables. The joint distribution for the phase and magnitude of each mode can be obtained from (3.1.16) and is given by[10]

$$P\left(\left|f_{\vec{k}}\right|, \arg f_{\vec{k}}\right) = \frac{2\left|f_{\vec{k}}\right|}{S_{\vec{k}}} \exp\left[-\frac{\left|f_{\vec{k}}\right|^2}{S_{\vec{k}}}\right] \frac{1}{2\pi}. \tag{3.1.21}$$

This distribution is illustrated in Figure 3.1. Given the distribution in (3.1.21), the mean of $\left|f_{\vec{k}}\right|$ is obtained from the properties of the Rayleigh distribution above and is

$$\left\langle |f_{\vec{k}}| \right\rangle^2 = \frac{\pi}{4} S_{\vec{k}}. \tag{3.1.22}$$

---

[10] Compare to [7, equation (2.6)] and note their apparent error in omitting the factor of $\frac{1}{2}$ that must be associated with $S_{\vec{k}} = \left\langle |f_{\vec{k}}|^2 \right\rangle$.

3.2 Random Fields and Cosmology

In this section I will briefly make the connection between the notation and expressions introduced above and the nomenclature used here and in the cosmology literature to describe cosmological perturbations. Any number of fields are used to describe the distribution of the contents and structure of the universe. In all cases one decomposes any quantity into its homogeneous background value and a perturbation away from it. For example, there is the density contrast, $\delta = (\rho - \rho_{\mathrm{b}}) / \rho_{\mathrm{b}}$, the metric perturbation, Φ, etc.. Often the perturbations are described by fields that are introduced in such a way as to be dimensionless, as in the density contrast above, but this is not always the case. These fields will all be defined precisely later but for now it suffices to say that they are all taken to have been Gaussian random fields at the earliest stages of structure formation. Historically this choice was motivated primarily for its mathematical simplicity but was also justified with some heuristic arguments stemming from the galaxy number-density distribution on the sky and the central limit theorem. Today, cosmological inflation provides a somewhat rigorous mechanism for explaining why quantities like the metric perturbation could be Gaussian random fields.[11] The fields are also taken to be homogeneous and isotropic (in the random-fields sense of the words). The imposition of these properties is motivated by the observational evidence that this is how our universe is — distant objects being isotropically distributed on our sky and apparently devoid of any significant long-range correlations. For a co-ordinate system for the fields, the most convenient choice is to use the FRW co-moving (with respect to the background) co-ordinates which makes the wave-number, $\vec{k}$, appearing in a field's Fourier transform the "conformal" wavenumber, in units of radians per co-ordinate interval rather than radians per physical distance. The means of the perturbation fields are all 0. Finally, we will consider their domain of definition to be infinite. This is consistent with our standing assumption of a $k = 0$ FRW universe.

Many authors take the fields to be confined to a finite box. This makes the fields' Fourier transforms discrete and in the context of a $k = 0$ FRW universe, which has infinite volume, is interpreted as simply imposing periodicity on the fields. The motivation for formulating the problem in a box is that the resulting discrete spectrum is amenable to numerical analysis on a digital computer.[12] If the box is larger than any visible volume (or just larger than the maximum correlation length), the periodicity has no impact on physical observables.[13] In this document, there is no advantage to this approach.

The remainder of this section will deal exclusively with homogeneous isotropic Gaussian random fields with means of 0 in three dimensions.

---

[11] Whether the predicted fluctuations are or are not Gaussian depends on the specifics of the inflation model. Inflation can be tuned to produce either. See, for example, [80].

[12] See, for example, [14].

[13] See [74, Section 26].

3.2.1 Power Spectrum

Following the discussion above, for homogeneous isotropic Gaussian random fields with 0 mean, very little information beyond just that is required in order to fully characterize them. In particular the spectral density function, S_k, as defined in (3.1.15), is sufficient.[14] In cosmology literature, this issue is somewhat confused. All authors agree that some sort of object which they generically refer to as the "power spectrum" is sufficient to characterize such a random field. What it is that each author means by "power spectrum," however, is often not clear and when it is they are rarely in agreement. Before investigating this further, it is useful to perform some manipulations of (3.1.14). For an isotropic field, $S_{\vec{k}} = S_k$ and we can perform the angular integrations in (3.1.14), pull a factor of k^{-1} out of the integrand, and write

$$\begin{aligned}\langle f^2(r)\rangle &= \frac{1}{(2\pi)^3}\int S_k \,\mathrm{d}^3k \\ &= \frac{4\pi}{(2\pi)^3}\int S_k k^2 \,\mathrm{d}k \\ &= \int \left(\frac{k^3}{2\pi^2}S_k\right)\frac{\mathrm{d}k}{k}. \end{aligned} \tag{3.2.1}$$

We are now in a position to compare expressions. In this document, $\mathcal{P}(k)$ will be used to denote the power spectrum and when confusion might arise, the particular field for which it is the spectrum will be indicated with a subscript, for example $\mathcal{P}_f(k)$.

Table 3.1 shows a comparison of the definitions of power spectrum used by some authors. When $-k$ appears on the right-hand side of one of the entries, it indicates a reversal of the definitions of the forward and inverse Fourier transforms used in that document compared to this. In general this is a significant difference but for real-valued fields, as are usually investigated in cosmology, it is not since $S_{-k} = S_k$ for such fields. Factors of 2π are, likewise, accumulated due to differences in Fourier transform definitions and are very significant: a factor of 2π is nearly an order of magnitude; six factors of 2π, as separates the definitions of some authors, introduces nearly 5 orders of magnitude discrepancy in the definition of an object known experimentally to within 10%. Readers must be very careful of this when mixing expressions from different authors. V is the volume of the finite cube whose boundaries are used to discretize the Fourier transform in the given document and is not, strictly speaking, significant since in this document the box is infinite in size and the role of the volume factor is played by the δ-function in (3.1.10).[15] There are, however, some discrepancies in the power of V that appears in the various definitions. Overall, the texts by Coles and Lucchin, [20], and Liddle and

---

[14] Looking at (3.1.11) one sees that the two-point covariance function $B(r)$ can be obtained from S_k through a Fourier transform. $B(r)$, in turn, allows the construction of any of the m-point joint probability distributions which are the defining properties of the field. Even for homogeneous anisotropic fields, $S_{\vec{k}}$ is sufficient but this is not true for inhomogeneous fields (isotropic or otherwise) where the definition of $S_{\vec{k}}$ is ambiguous.

[15] See, for example, the expressions in Footnote 24 in Chapter 3. In any case, no physically relevant quantity can depend on V so it should not appear in any "final results."

Author(s)	Definition
Bardeen, Bond, Kaiser and Szalay [7]	$\mathcal{P}(k) = (2\pi)^{-3} S_k$
Bergström and Goobar [11, Section 12.3]	$\mathcal{P}(k) = S_k$
Coles and Lucchin [20, Section 13.2]	$\mathcal{P}(k) = S_k$
Fan and Bardeen [23]	$\mathcal{P}(k) = S_{-k}$
Kolb and Turner [48, Section 9.2.2]	$\mathcal{P}(k) = V^{-1} S_{-k}$
Liddle and Lyth [55, equation (4.15)]	$\mathcal{P}(k) = V \frac{k^3}{2\pi^2} S_k$
Liddle and Lyth [55, equation (4.17)]	$\mathcal{P}(k) = V S_k$
Mukhanov, Feldman, Brandenberger [69, Section 12]	$\mathcal{P}(k) = \lvert\delta(\eta, k)\rvert^2 = \frac{k^3}{2\pi^2} S_k$
Narlikar [70, Section 7.6.2]	$\mathcal{P}(k) = F(k) = (2\pi)^3 S_k$
Padmanabhan [72, Section 5.3]	$\mathcal{P}(k) = S_k$
Peebles [74, Section 41]	$\mathcal{P}(k) = V^{-1} S_{-k}$

Table 3.1: Some definitions of "power spectrum" found in the wild in terms of the equivalent quantities used in this document. Coles and Lucchin are the only authors I have found who point out that their power spectrum is more correctly called the spectral density function. For the second definition provided by Liddle and Lyth, they say this is "used only for the matter density perturbations" and in a later footnote (page 75) explain that this is the "older definition of the spectrum." Mukhanov et al. in [69] do not actually use the term "power spectrum" but introduce $|\delta(\eta, k)|^2$ which is equivalent and described as "a measure of the square of the amplitude of fluctuations [...] at co-moving wavelength $1/k$." Likewise, Narlikar in [70] does not use the term "power spectrum" but introduces $F(k)$ which he calls simply the "spectrum."

Lyth, [55], give the most technically coherent treatments of cosmological random fields in general and the power spectrum in particular.

In summary there are essentially two conventions for the meaning of "power spectrum":

1. "power spectrum" is a synonym for spectral density function, S_k, modulo factors of 2π, V, and i.
2. "power spectrum" is the contribution to $\langle f^2 \rangle$ per logarithmic interval in wavenumber k.

For the purposes of this document, I will use the second of the two choices above. In other words, for a homogeneous isotropic random field $f(r)$, the power spectrum, $\mathcal{P}(k)$, is defined by

$$\int \mathcal{P}(k)\,\mathrm{d}\ln k = \int \mathcal{P}(k)\frac{\mathrm{d}k}{k} \equiv \langle f^2 \rangle, \tag{3.2.2}$$

which, from (3.2.1), means that

$$\mathcal{P}(k) \equiv \frac{k^3}{2\pi^2} S_k. \tag{3.2.3}$$

Some Other Relationships

We can obtain some other useful relationships from the expression for the power spectrum in (3.2.3). Combining (3.1.11), (3.1.15) and (3.2.3) gives

$$\sqrt{\frac{\pi}{2}}\frac{1}{k^2}\mathcal{P}(k) = \sqrt{\frac{2}{\pi}}\int rB(r)\sin(kr)\,\mathrm{d}r. \tag{3.2.4}$$

We see that $\sqrt{\frac{\pi}{2}}\frac{1}{k^2}\mathcal{P}(k)$ is the Fourier sine transform of $rB(r)$ where the Fourier sine transform and its inverse are given by[16]

$$F(y) = \sqrt{\frac{2}{\pi}}\int_0^\infty f(x)\sin(xy)\,\mathrm{d}x \tag{3.2.5a}$$

$$f(y) = \sqrt{\frac{2}{\pi}}\int_0^\infty F(x)\sin(xy)\,\mathrm{d}y. \tag{3.2.5b}$$

That fact isn't too interesting, but using (3.2.5b) to invert (3.2.4) gives[17]

$$B(r) = \int_0^\infty \mathcal{P}(k)\frac{\sin(kr)}{kr}\frac{\mathrm{d}k}{k} \tag{3.2.6}$$

which is a much cleaner relationship. It is important to note that when using the definition of power spectrum chosen here, it and the field's two-point covariance function

---

[16] See [15, equation (4.68), Section 4.4.2.1].

[17] Identical to [55, equation (4.20)] and [69, equation (12.13)] with a reminder that for a random homogeneous field with 0 mean, the covariance function is equal to the correlation function, i.e. $B(r) = \xi(r)$.

(or correlation function — remember we're dealing with fields of 0 mean) are not merely the Fourier transforms of each other as is frequently claimed by authors of cosmology texts.[18] When an author makes that claim, it is a sure sign that they are using the other convention, the first listed above, for the meaning of power spectrum.

Another very important relationship is obtained by substituting (3.2.3) into (3.1.15) and then into (3.1.10) to get[19]

$$\left\langle f_{\vec{k}} f^*_{\vec{k}'} \right\rangle = \delta^{(3)}\left(\vec{k} - \vec{k}'\right) \frac{2\pi^2}{k^3} \mathcal{P}(k), \tag{3.2.7}$$

giving the field's two-point spectral correlation function in terms of its power spectrum.

Although, throughout this document, the power spectrum will be used as the object of choice for describing the statistical properties of cosmological random fields, this is done with a reminder that there are almost as many definitions of "power spectrum" in the literature as there are places where its definition is written out. Unless the power spectrum is carefully defined, it really should be the spectral density function that is taken as the description of a field's statistical properties when moving from one author's work to that of another.

3.2.2 Spectral Index

Often the power spectrum is approximated by assuming it to be proportional to some power of k. This power-law structure for the power spectrum is parameterized by a quantity called the "spectral index," denoted n_s, but here again authors have used many different ways to turn the exponent of k into a parameter. Most authors choose to define the spectral index so that their definition of power spectrum and their definition of spectral index are related to one another by $\mathcal{P}(k) \propto k^{n_s}$. There are, however, exceptions to this rule and this document is one of them!

The particular definition of spectral index that will be used in this document is the same as that used in [48, Section 9.4], [72, Section 4.7] and [87]. In particular, for any power spectrum, $\mathcal{P}(k)$, as defined in (3.2.3),

$$n_s(k) = 1 + \frac{\mathrm{d}}{\mathrm{d}\ln k} \ln \mathcal{P}(k). \tag{3.2.8}$$

For a pure power-law spectrum, $n_s(k)$ is independent of k and related to the power spectrum by

$$\mathcal{P}(k) \propto k^{n_s - 1}. \tag{3.2.9}$$

For the curious, the two standard meanings for the term power spectrum differ by k^3. So when an author is using the spectral density function as the power spectrum, and using a n_s that is related to it by $\mathcal{P}(k) \propto k^{n_s}$, their spectral index is the spectral index used in this document minus 4. For example, the "scale-invariant spectrum" that is described by $n_s = 1$ in this document would be described by $n_s = -3$ for those authors.

---

[18] See [74, equation (41.10)].

[19] Identical to [55, equation (4.26)].

"Scale-Invariance"?!

The $n_s = 1$ power spectrum mentioned above is sometimes called the "scale-invariant" spectrum. This term will be avoided in this document because this is yet another aspect of cosmological random fields for which there is more than one common meaning. In the particular case of the $n_s = 1$ spectrum above, the idea of scale invariance can be understood by recognizing that the operation of rescaling the spatial co-ordinates means a rescaling of k which corresponds to a translation on a logarithmic scale, and so from (3.2.2) a "scale-invariant power spectrum" — one that is left unchanged by a co-ordinate rescaling — is one for which $\mathcal{P}(k)$ is independent of k.

One will also see authors describing a random field for which $S_k \propto k$ as having a "scale-invariant" spectrum.[20] Confusion ensues when the same authors who do this choose to define the power spectrum as $\mathcal{P}(k) = S_k$ so that their "scale-invariant" spectrum is *also* described by their $n_s = 1$. Such a field has $n_s = 5$ using the conventions in this document. The reason this spectrum is called "scale-invariant" is that this is the spectrum of matter density contrast fluctuations one sees in an expanding dust-filled Robertson-Walker universe when the primordial density fluctuations all have the same amplitude as they enter the horizon. This spectrum is also called the Harrison-Zeldovich spectrum.[21]

Curiously, and even more confusingly, in the context of cosmological perturbations these two types of random fields are closely related. This will all be discussed in detail later but for now imagine one introduces a perturbation to the geometry of space-time parameterized by a field Φ that is interpreted as a generalized Newtonian gravitational potential. At the same time imagine introducing a perturbation to the matter density parameterized by a random field δ. It will be seen below that for appropriately defined Φ and δ they are related by $\delta \propto \nabla^2\Phi$ which means their Fourier transforms are related by $\delta_k \propto k^2\Phi_k$ and so their power spectra are related by $\mathcal{P}_\delta(k) = k^4\mathcal{P}_\Phi(k)$. Finally, this means that $n_{s\delta} = n_{s\Phi} + 4$ or in other words for a single realization of the cosmological perturbations both Φ and δ can be "scale-invariant" simultaneously where in the case of the former we have $n_s = 1$ and in the case of the latter we have $n_s = 5$. It is then left as an exercise to the reader to keep track of which is which and for what reason.

If this wasn't enough confusion, under certain approximations, the relationship between δ and Φ reduces to $\delta \propto \Phi$ which gives $n_{s\delta} = n_{s\Phi}$ and so once again the two can be "scale-invariant" simultaneously but now for entirely different reasons.

3.3 Units

One last comment that should be made before moving on is with regard to the units carried by the quantities introduced thus far. If the co-ordinates, $\vec{r}$, carry units of [co-ordinate units] and the field, $f(\vec{r})$, carries units of [field units] then Table 3.2 shows the units carried by other things.

---

[20] See [11, Section 12.3] and [20, Section 13.4].

[21] See [20, Section 13.4], [33], [99].

Quantity	Units
$\vec{r}$	[co-ordinate units]
$f(\vec{r})$	[field units]
$f_{\vec{k}}$	[field units][co-ordinate units]^n
$B(\vec{r})$	[field units]^2
$S_{\vec{k}}$, $\left\langle \lvert f_{\vec{k}} \rvert^2 \right\rangle$	[field units]^2[co-ordinate units]^{2n}
$\mathcal{P}(k)$	[field units]^2

Table 3.2: The units carried by the various quantities used to describe random fields in terms of the units carried by the co-ordinates and the field itself. The parameter n is the number of dimensions in the space in which the field is defined.

3.4 Properties of Gaussian Random Fields

In this section we will study some of the properties of Gaussian random fields. We will continue from above considering homogeneous isotropic Gaussian random fields with means of 0 in three dimensions. We will change notation slightly, though: from now on $\vec{x}$ will be used to denote field co-ordinates instead of $\vec{r}$ since r will now be used to indicate a smoothing radius.

3.4.1 Smoothing

Imagine that we wish to obtain the spatial average of f within a region of radius r (to be measured in a manner described below) centred on $\vec{x}_0$. Specifically, we take each realization of the field and compute the average value of the field over some volume; an ensemble of such averages is then constructed from the ensemble of realizations of the random field. The result is a random variable whose distribution is the distribution of averages over the specified volume. By repeating the process for each possible choice of volume centre a new, smoothed, random field is constructed. The spatial average will be performed according to a weighting function, $W(\vec{x}; r)$, often called the "window function" or "window profile." The weighting function has dimensions of [volume]$^{-1}$ and, in this document, is normalized so that its integral, or "volume," is 1.[22] The mean value of f averaged over a region of characteristic radius r centred at $\vec{x}_0$ is

$$\overline{f}(\vec{x}_0; r) = \int_{\vec{x}} W(\vec{x} - \vec{x}_0; r)\, f(\vec{x})\, \mathrm{d}^3x. \tag{3.4.1}$$

We will call this the smoothed field. The smoothing procedure is linear so the smoothed field is also a Gaussian random field and is easily shown to be homogeneous and isotropic

---

[22] Some authors, for example [55, Section 4.3.3], do not normalize $W(\vec{x}; r)$ which results in them carrying around a factor of V, the window function's "volume"; not to be confused with V, the volume of the finite unit cell in which they perform Fourier transforms. It is often left as an exercise for the reader to remember which V is which.

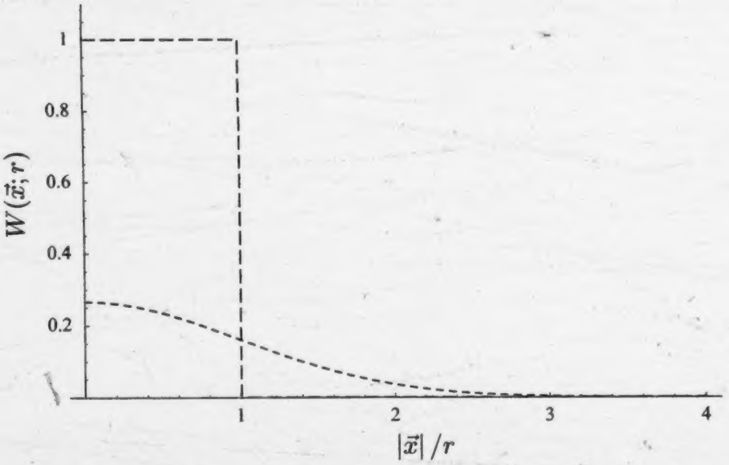


Figure 3.2: The radial behaviour of the two standard window functions. The short dashes illustrate the Gaussian window while the long dashes show the top-hat window. The the vertical axis is normalized to the top-hat window at the origin. Note that the apparent discrepancy in normalization is due to the omission of the angular contribution to the volume element — the volume integral of each function really is 1.

like the underlying field. In the analysis that follows, the window function is left unspecified but there are two standard choices and they are the "top-hat" window given by

$$W(\vec{x};r) = \begin{cases} \frac{3}{4\pi}r^{-3} & \text{for } |\vec{x}| \leq r, \\ 0 & \text{for } |\vec{x}| > r \end{cases} \tag{3.4.2}$$

and the "Gaussian" window given by

$$W(\vec{x};r) = \frac{1}{\sqrt{2\pi}^3 r^3} \exp\left(-\frac{1}{2}\frac{|\vec{x}|^2}{r^2}\right). \tag{3.4.3}$$

It is important to realize that we are using a radius as the parameter for the window functions, not a diameter. Throughout this document, terms like "smoothing length," "smoothing size," "characteristic size," etc., will be used interchangeably and in all cases it is the parameter r that is meant. The radial dependence of the two window functions above can be compared in Figure 3.2.

The interpretation of the top-hat window is the most straight-forward since an average computed using it is clearly simply the unweighted mean value within a ball of radius r. The reason one considers the conceptually less transparent Gaussian window is that its Fourier transform is much more localized in frequency space than the top-hat window's. How the smoothing procedure relates to a field's spectral behaviour will be investigated in detail below but it is helpful, here, to look at a simple example. The effect of smoothing a random field with both top-hat and Gaussian windows is shown in Figure 3.3. The field shown is one realization of a $n_s = 1$ random field. As can be seen, the two types of smoothed fields are broadly similar although it is clear the top-hat window retains more small-scale structure for the same smoothing radius. This phenomenon is easily understood: because the top-hat window has sharp edges, a large fluctuation can sit just outside it and then just inside it when the window is slid over a short distance; this can't happen with the Gaussian window since a large fluctuation will be gradually felt more and more strongly as the window is moved closer to it. Because of this phenomenon, it will be found that the Gaussian window function allows us to work with fields whose spectral behaviour would otherwise give divergent results due to high-frequency modes when smoothed with the top-hat window.

For both window functions listed above, $W(\vec{x};r)$ is invariant under reflection through the origin,

$$W(\vec{x};r) = W(-\vec{x};r)\,. \tag{3.4.4}$$

All window functions considered in this document will possess this property which allows the mean density from (3.4.1) to be rewritten in the equivalent but more useful form[23]

$$\overline{f}(\vec{x}_0;r) = \int_{\vec{x}} W(\vec{x}_0 - \vec{x};r)\, f(\vec{x})\, \mathrm{d}^3x. \tag{3.4.5}$$

Within the region centred on $\vec{x}_0$, we can describe the field in terms of its departure from the local spatial average. First of all we introduce a set of co-ordinates, $\vec{x}'$, for use within the region. The origin of these co-ordinates is set to the centre of the window function, $\vec{x}_0$,

$$\vec{x} = \vec{x}_0 + \vec{x}'. \tag{3.4.6}$$

We can then write

$$f(\vec{x}_0 + \vec{x}') = \overline{f}(\vec{x}_0;r) + \tilde{f}(\vec{x}') \tag{3.4.7}$$

which defines $\tilde{f}$, the "local" field — the underlying field represented as a perturbation away from the local spatial average.

It should be noted that while

$$\langle f \rangle = 0 \tag{3.4.8}$$

is true by definition, it is also true that

$$\langle \overline{f} \rangle = 0 \tag{3.4.9}$$

and

$$\left\langle \tilde{f} \right\rangle = 0, \tag{3.4.10}$$

and also that

$$\left\langle W(\vec{x}';r)\, \tilde{f}(\vec{x}') \right\rangle_{\vec{x}'} = 0. \tag{3.4.11}$$

---

[23] This can always be done anyway, regardless of the parity-invariance or otherwise of the window function. It's only necessary to define the window function to be the mirror image of whatever it is you actually want.

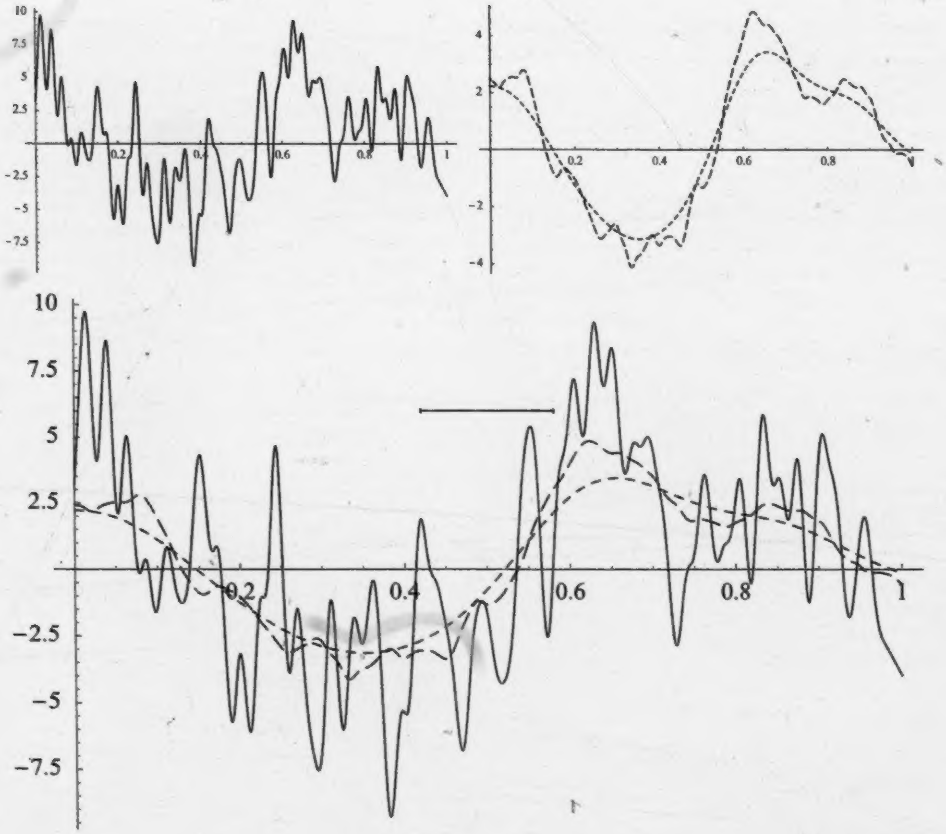


Figure 3.3: Smoothing a realization of a random field in one dimension. At the top-left, the random field. At the top-right, the random field after being smoothed using a top-hat window (long dashes) and a Gaussian window (short dashes). At bottom, all three curves superimposed. The scale indicator above the curves in the bottom image shows twice the smoothing radius — the "width" of the window functions in one dimension.

Throughout this document, a subscripted quantity attached to the averaging brackets indicates that the procedure is to be performed over choices of that given parameter. If no parameter is indicated then the average is to be performed over the ensemble of realizations of the random variable in the brackets.

We'll define the Fourier transforms of the random field, the window function, and the smoothed field (as a function of $\vec{x}_0$) in the usual way, namely[24]

$$f(\vec{x}) = \frac{1}{\sqrt{2\pi}^3} \int f_{\vec{k}} \mathrm{e}^{\mathrm{i}\vec{k}\cdot\vec{x}} \,\mathrm{d}^3 k \tag{3.4.12}$$

$$W(\vec{x};r) = \frac{1}{\sqrt{2\pi}^3} \int W_{\vec{k};r} \mathrm{e}^{\mathrm{i}\vec{k}\cdot\vec{x}} \,\mathrm{d}^3 k \tag{3.4.13}$$

$$\overline{f}(\vec{x}_0;r) = \frac{1}{\sqrt{2\pi}^3} \int \overline{f}_{\vec{k};r} \mathrm{e}^{\mathrm{i}\vec{k}\cdot\vec{x}_0} \,\mathrm{d}^3 k. \tag{3.4.14}$$

The convolution theorem states that if given a convolution product,

$$(f * g)(x) = \int_{-\infty}^{+\infty} f(y) g(x-y) \,\mathrm{d}y, \tag{3.4.15}$$

then the Fourier transform of the convolution is (to within factors of $\sqrt{2\pi}$) equal to the product of the Fourier transforms of the separate factors. More precisely,[25]

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (f * g)(x) \mathrm{e}^{\mathrm{i}xy} \,\mathrm{d}x = \sqrt{2\pi} F(y) G(y). \tag{3.4.16}$$

The convolution theorem can now be seen to be the motivation for having rewritten the original definition of the averaging procedure in (3.4.1) as (3.4.5). In doing so, we can now use the convolution theorem to perform the averaging procedure in frequency space. Applying the convolution theorem to (3.4.5) gives

$$\overline{f}_{\vec{k};r} = \sqrt{2\pi}^3 W_{\vec{k};r} f_{\vec{k}} \tag{3.4.17}$$

as the relationship between the Fourier components of the smoothed field and those of the underlying field. From this and the relationship in (3.2.7) between a homogeneous isotropic field's two-point spectral correlation function and its power spectrum, we can obtain a relationship between the power spectrum of the field and its smoothed counterpart, namely

$$\mathcal{P}_{\overline{f}}(k) = (2\pi)^3 W_{k;r}^2 \mathcal{P}_f(k). \tag{3.4.18}$$

---

[24] In many texts, the universe is considered to be a large box. This results in a discrete spectrum and the integral definitions of the Fourier transform become sums. If one wishes to do so, the equivalent definitions for the Fourier transform and its inverse are: $f(x) = \frac{1}{\sqrt{2\pi}^3} \sum_{\vec{k}=-\infty}^{+\infty} f_{\vec{k}} \mathrm{e}^{\mathrm{i}\vec{k}\cdot\vec{x}}$; $f_{\vec{k}} = \frac{1}{\sqrt{2\pi}^3 V} \int_V f(\vec{x}) \mathrm{e}^{-\mathrm{i}\vec{k}\cdot\vec{x}} \,\mathrm{d}^3 x$; $\vec{k} = \frac{2\pi}{V^{1/3}} \vec{n}$; $n_x, n_y, n_z \in$ Integers.

[25] See [15, Section 4.4.2.1].

In general, if two fields, f and g, are related by $g_{\vec{k}} = X_{\vec{k}} f_{\vec{k}}$ then

$$\mathcal{P}_g(k) = X_k^2 \mathcal{P}_f(k). \tag{3.4.19}$$

The Fourier transform of the top-hat window function is

$$W_{\vec{k};r} = \frac{1}{\sqrt{2\pi}^3} \frac{3}{k^3 r^3} (\sin kr - kr \cos kr) \tag{3.4.20}$$

$$= \frac{3}{\sqrt{2\pi}^3 kr} \mathrm{j}_1(kr) \tag{3.4.21}$$

where j_1 is the spherical Bessel function of order 1,[26]

$$\mathrm{j}_1(x) = \frac{1}{x^2} (\sin x - x \cos x). \tag{3.4.22}$$

Note that the Fourier transform is real-valued. In general, $W(\vec{x};r)$'s property of being invariant under $\vec{x} \to -\vec{x}$ means that its Fourier transform will always be real-valued. It is useful to comment on the Bessel function's asymptotic behaviour. For small x, we have

$$\mathrm{j}_1(x \ll 1) = \frac{1}{x^2} \left(x - \frac{x^3}{3!} + \cdots - x + \frac{x^3}{2!} + \cdots \right) \approx \frac{x}{3}, \tag{3.4.23}$$

while for large x

$$\mathrm{j}_1(x \gg 1) \approx -\frac{\cos x}{x}. \tag{3.4.24}$$

The short and long wavelength asymptotic limits of the Fourier transform of the top-hat window function are readily obtained from these results and for small kr,

$$W_{\vec{k};r} \approx \frac{1}{\sqrt{2\pi}^3} \tag{3.4.25}$$

while for large kr,

$$W_{\vec{k};r} \approx \frac{-3}{\sqrt{2\pi}^3} \frac{\cos kr}{k^2 r^2}. \tag{3.4.26}$$

The Fourier transform of the Gaussian window is[27]

$$W_{\vec{k};r} = \frac{1}{\sqrt{2\pi}^3} \exp\left(-\frac{1}{2} k^2 r^2 \right). \tag{3.4.27}$$

For small kr this is also $\approx 1/\sqrt{2\pi}^3$ while for large kr it drops off exponentially rather than polynomially. This is the key to the Gaussian window being able to handle less

---

[26] See [3, equation (11.154)]. Note that this definition appears to differ from that given in [15, Section 3.3.1.3.4]. I cannot explain the discrepancy but since this is the more convenient form, it will be the one I use.

[27] Obtaining this requires the Fourier sine transformation found in [15, Section 4.4.2.2].

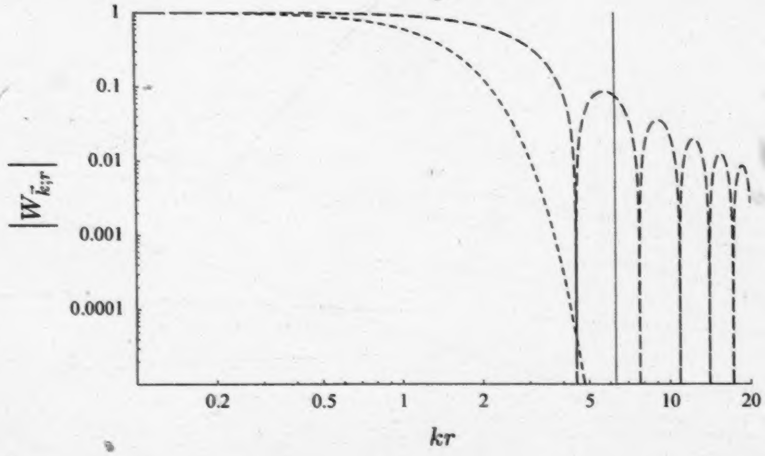


Figure 3.4: The frequency dependence of the two standard window functions. The short dashes show the Gaussian window, while the long dashes show the top-hat window. The horizontal axis has the reciprocal smoothing radius indicated at $k = 2\pi/r$. The vertical axis is normalized to the top-hat window at the origin.

well-behaved fluctuation spectra. A plot of the frequency-space representations of the two window functions is shown in Figure 3.4.

3.4.2 The Properties of $\overline{f}$, Part I

In this section we will investigate the statistical properties of $\overline{f}$. We already have its mean in (3.4.9). Its variance can be found by substituting (3.4.17) into (3.4.14) which gives an expression for the smoothed field in terms of the Fourier decomposition of the background,

$$\overline{f}_{\vec{x}_0;r} = \int W_{\vec{k};r} f_{\vec{k}} \mathrm{e}^{\mathrm{i}\vec{k}\cdot\vec{x}_0}\, \mathrm{d}^3k. \tag{3.4.28}$$

From this and using (3.1.10), the mean square is

$$\begin{aligned}
\left\langle \overline{f}^2(\vec{x}_0;r) \right\rangle &= \left\langle \int W_{\vec{k};r} W_{\vec{k}';r} f_{\vec{k}} f^*_{\vec{k}'} \mathrm{e}^{\mathrm{i}(\vec{k}-\vec{k}')\cdot\vec{x}_0}\, \mathrm{d}^3k\, \mathrm{d}^3k' \right\rangle \\
&= \int W_{\vec{k};r} W_{\vec{k}';r} \left\langle f_{\vec{k}} f^*_{\vec{k}'} \right\rangle \mathrm{e}^{\mathrm{i}(\vec{k}-\vec{k}')\cdot\vec{x}_0}\, \mathrm{d}^3k\, \mathrm{d}^3k' \qquad (3.4.29) \\
&= \int W_{\vec{k};r} W_{\vec{k}';r} S_{\vec{k}} \delta(\vec{k}' - \vec{k}) \mathrm{e}^{\mathrm{i}(\vec{k}-\vec{k}')\cdot\vec{x}_0}\, \mathrm{d}^3k\, \mathrm{d}^3k'
\end{aligned}$$

$$\left\langle \overline{f}^2(\vec{x}_0;r) \right\rangle = \int W^2_{\vec{k};r} S_{\vec{k}}\, \mathrm{d}^3k. \tag{3.4.30}$$

Remember that $\overline{f}(\vec{x}_0; r)$ is a Gaussian random variable so the first two moments are all that are needed in order to fully characterize it. Because its mean is 0, its distribution is given by

$$P(\overline{f}; r) = \frac{1}{\sqrt{2\pi \left\langle \overline{f}^2 \right\rangle}} \exp\left[-\frac{1}{2} \frac{\overline{f}^2}{\left\langle \overline{f}^2 \right\rangle} \right] \tag{3.4.31}$$

which is the probability of observing a given region of size r to have an average field value between $\overline{f}$ and $\overline{f} + \mathrm{d}\overline{f}$.

Let's investigate the variance more closely. Since the field is isotropic, $S_{\vec{k}} = S_k$ which allows us to perform the angular integrations in (3.4.30) giving[28]

$$\begin{aligned} \left\langle \overline{f}^2 \right\rangle &= \int W^2_{\vec{k};r} S_k k^2 \,\mathrm{d}k \,\mathrm{d}^2\Omega. \\ &= 4\pi \int_{k=0}^{\infty} W^2_{k;r} S_k k^2 \,\mathrm{d}k \\ &= (2\pi)^3 \int_{k=0}^{\infty} W^2_{k;r} \mathcal{P}(k) \frac{\mathrm{d}k}{k}. \end{aligned} \tag{3.4.32}$$

This result could have been obtained without first deriving (3.4.30). Had we simply substituted (3.4.18) into (3.2.2) the result would have immediately been (3.4.32). Given the confusion that exists in the literature surrounding these sorts of relationships, it is useful to verify for ourselves that the expressions we are using are all internally consistent by checking that the same result can be obtained by combining the relationships in more than one way. Continuing, let's assume a standard power-law spectrum for the underlying field such that

$$\mathcal{P}(k) = \sigma^2 k^{n_s - 1} \tag{3.4.33}$$

where n_s is the spectral index and σ^2 is some constant for setting the amplitude of the power spectrum. With the top-hat window function, the variance in (3.4.32) becomes

$$\begin{aligned} \left\langle \overline{f}^2 \right\rangle &= (2\pi)^3 \sigma^2 \int_{k=0}^{\infty} \left(\frac{3}{\sqrt{2\pi}^3 kr} \mathrm{j}_1(kr) \right)^2 k^{n_s - 2} \,\mathrm{d}k \\ &= \frac{\sigma^2}{r^{n_s - 1}} 9 \int_{x=0}^{\infty} x^{n_s - 8} \left(\sin x - x \cos x \right)^2 \,\mathrm{d}x. \end{aligned} \tag{3.4.34}$$

---

[28] This result differs from all manner of expressions found in the literature. In [63], for example, in their equation (33) they have a prefactor of $\frac{1}{2\pi^2}$ rather than the 4π that appears in the second line here. In order to explain the discrepancy it would be necessary to examine the form they are using for the Fourier transform for which they provide, immediately subsequently, two conflicting definitions. For this reason I cannot comment on such differences beyond pointing them out.

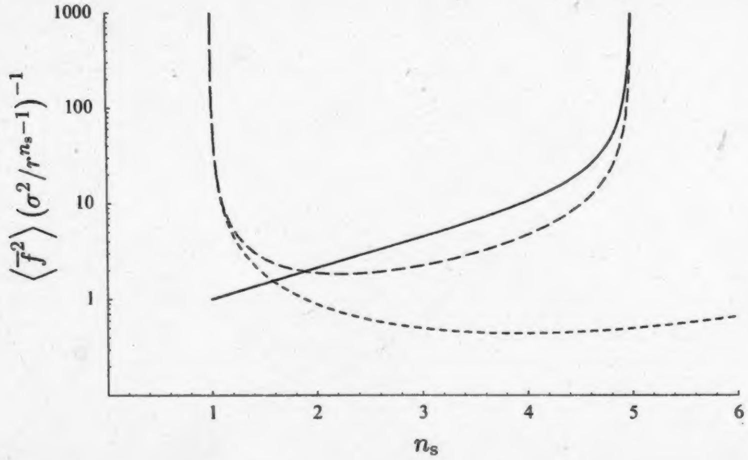


Figure 3.5: The behaviour of $\left\langle \overline{f}^2 \right\rangle \left(\sigma^2/r^{n_s-1}\right)^{-1}$ as a function of the spectral index n_s for two choices of window function. The large dashes correspond to the top-hat window, the short dashes to the Gaussian window. The solid curve gives the ratio of the two results.

Repeating the integration of (3.4.32) for the Gaussian window gives

$$\begin{aligned}\left\langle \overline{f}^2 \right\rangle &= (2\pi)^3\sigma^2 \int_{k=0}^{\infty} \left[\frac{1}{\sqrt{2\pi}^3} \exp\left(-\frac{1}{2}k^2r^2\right)\right]^2 k^{n_s-2}\,\mathrm{d}k \\ &= \frac{\sigma^2}{r^{n_s-1}} \int_{x=0}^{\infty} x^{n_s-2} \exp\left(-x^2\right)\,\mathrm{d}x\end{aligned}$$

and for $n_s \geq 1$ this is

$$\left\langle \overline{f}^2 \right\rangle = \frac{\sigma^2}{r^{n_s-1}} \frac{1}{2} \Gamma\left(\frac{n_s-1}{2}\right). \tag{3.4.35}$$

A comparison of the effects of the choice of window function is shown in Figure 3.5 which, for both (3.4.34) and (3.4.35) is a plot of $\left\langle \overline{f}^2 \right\rangle \left(\sigma^2/r^{n_s-1}\right)^{-1}$ as a function of the spectral index n_s — in other words, that part of $\left\langle \overline{f}^2 \right\rangle$ which depends only the choice of window function. As can be seen, as n_s increases from a value of 1 the choice of window function becomes more and more significant. The differences between them become most acute for $n_s \geq 5$ where the smoothed density contrast computed with the top-hat window diverges while the Gaussian window still gives well-behaved results. This is the result of the Gaussian window's suppressed sensitivity to high-frequency modes. In fact the smoothed density contrast computed using a Gaussian window, although increasing without bound for large n_s, remains finite for *all* spectra with $n_s > 1$.

The problem with a $n_s = 1$ field is that its spectrum contains too much power at long wavelengths: $\mathcal{P}(k)$ does not fall off rapidly enough (indeed not at all) as $k \to 0$.

Of course, this problem exists regardless of the choice of window function since the two differ only in their sensitivity to short wavelength fluctuations, not long.

The problems encountered with the $n_s = 1$ spectrum would be neither here nor there except that the observational evidence points to this, or even $n_s < 1$, being exactly the sort of fluctuation spectrum the perturbation fields of our universe started out with, so we really do need to take a closer look at this problem. As stated above, it boils down to there being too much power in long wavelength modes. Examining (3.4.32), it is evident that the variance of a smoothed field will be finite if $\mathcal{P}(k) \to 0$ as $k \to 0$, even if it only does so rather abruptly at some small wavenumber. What about these long wavelengths? In the context of inflation and the inflaton field, the very longest wavelength perturbations are those that left the de Sitter horizon at the very start of the inflation epoch so their description requires knowledge of the physics at that time. We do not have this knowledge so it is impossible to say what the spectrum of inflaton perturbations really does as $k \to 0$. In other words, at least in the case of the inflaton field and its perturbations, there is no rigorous mechanism for solving the problem of the divergence of the variance of $\overline{f}$ for $n_s \leq 1$.

At the same time, the very longest wavelengths are invisible. In practise, one way or another, there is always a limit to the physical volume of space that can be observed; even a limit to the volume one is interested in theoretically. Any fluctuations whose wavelengths are much longer than the diameter of that volume cannot be distinguished from the (presumably) homogeneous background and can be considered to be part of it. If the radius of the largest volume we could be interested in is r_* and we assign perturbations with wavelengths longer than that to the homogeneous background of such regions, then the background itself becomes a random variable that differs from one region to the next. The uncertainty in the background can be incorporated into one's model or it can be thrown out. In this document we will discard this uncertainty by choosing r_*-size regions that look like the underlying field, i.e. for which the smoothed field on that scale equals 0,

$$\overline{f}(\vec{x}_0; r_*) = 0.$$

In order to proceed, we must examine the properties of perturbations away from the apparent homogeneous background inside a region of size r_*. The notation for doing so was introduced in (3.4.7) where the local field $\tilde{f}$ was defined. Let us now momentarily suspend our investigation of $\overline{f}$ and discuss $\tilde{f}$ some more.

3.4.3 The Properties of $\tilde{f}$

In this section we will examine the properties of the random field, $f(\vec{x})$, when described as a departure from the local spatial average, $\overline{f}(\vec{x}_0; r_*)$. The underlying field, f, is decomposed into two components: one component forms the spatial average or "local homogeneous background," $\overline{f}$, which was studied in the last section while the other component forms the fluctuations away from the local average, $\tilde{f}$, to be studied here. In frequency space, one can imagine a split being made between long wavelength fluctuations of the underlying field which go into forming the local apparent homogeneous background and short wavelengths which contribute to the local fluctuations. How this will occur can be seen by considering the window functions described in Section 3.4.1 and noting that they exhibit a localization about $\vec{k} = 0$ in frequency space. The consequence in (3.4.17) is that $\overline{f}_{\vec{k};r}$ will fall off for large k — only long wavelength fluctuations contribute to the background. In this way, it is our choice of the window function used to define an average that determines the split between wavelengths that contribute to the background and those that contribute to the foreground.

Let's do this precisely. We begin by rewriting (3.4.7) for $\tilde{f}$ as the departure of the underlying field from the local spatial average,

$$\tilde{f}(\vec{x}') = f(\vec{x}_0 + \vec{x}') - \overline{f}(\vec{x}_0; r_*). \tag{3.4.36}$$

Expressing both sides in terms of their Fourier transforms,

$$\frac{1}{\sqrt{2\pi}^3} \int \tilde{f}_{\vec{k}} e^{i\vec{k}\cdot\vec{x}'} \, d^3k = \frac{1}{\sqrt{2\pi}^3} \left[\int f_{\vec{k}} e^{i\vec{k}\cdot(\vec{x}_0 + \vec{x}')} \, d^3k - \int \overline{f}_{\vec{k};r_*} e^{i\vec{k}\cdot\vec{x}_0} \, d^3k \right], \tag{3.4.37}$$

and using (3.4.17) for $\overline{f}_{\vec{k};r_*}$ gives

$$\begin{aligned} \frac{1}{\sqrt{2\pi}^3} \int \tilde{f}_{\vec{k}} e^{i\vec{k}\cdot\vec{x}'} \, d^3k &= \frac{1}{\sqrt{2\pi}^3} \left[\int f_{\vec{k}} e^{i\vec{k}\cdot(\vec{x}_0 + \vec{x}')} \, d^3k - \int \sqrt{2\pi}^3 W_{\vec{k};r_*} f_{\vec{k}} e^{i\vec{k}\cdot\vec{x}_0} \, d^3k \right]. \\ &= \frac{1}{\sqrt{2\pi}^3} \int f_{\vec{k}} e^{i\vec{k}\cdot\vec{x}_0} \left[1 - \sqrt{2\pi}^3 W_{\vec{k};r_*} e^{-i\vec{k}\cdot\vec{x}'} \right] e^{i\vec{k}\cdot\vec{x}'} \, d^3k. \end{aligned} \tag{3.4.38}$$

At this point it would be helpful if the phase factor multiplying $W_{\vec{k};r_*}$ in the second term in the brackets in (3.4.38) could be dropped as everything remaining would then be independent of $\vec{x}'$. Qualitatively this operation can be justified by noting that wavelengths smaller than the region in question, $kr_* > 1$, should not contribute significantly to $\overline{f}(\vec{x}_0; r_*)$, the origin of the second term, since their positive and negative contributions within the region will tend to cancel. For this reason, introducing (or dropping) a phase factor that only becomes significant for $kr_* \gtrsim 1$ should not be a problem. This is, in fact, equivalent to the approximation

$$\overline{f}(\vec{x}_0 + \vec{x}'; r_*) \approx \overline{f}(\vec{x}_0; r_*) \tag{3.4.39}$$

which is not unreasonable since we have necessarily that $|\vec{x}'| \leq r_*$, with r_* being the scale over which the smoothing was done to obtain $\overline{f}$. Making the approximation turns (3.4.38) into

$$\frac{1}{\sqrt{2\pi}^3} \int \tilde{f}_{\vec{k}} e^{i\vec{k}\cdot\vec{x}'} \, d^3k = \frac{1}{\sqrt{2\pi}^3} \int f_{\vec{k}} \left(1 - \sqrt{2\pi}^3 W_{\vec{k};r_*} \right) e^{i\vec{k}\cdot\vec{x}} \, d^3k \tag{3.4.40}$$

and taking the inverse Fourier transform of both sides leaves the result

$$\tilde{f}_{\vec{k}} = f_{\vec{k}} \left(1 - \sqrt{2\pi}^3 W_{\vec{k};r_*} \right) e^{i\vec{k}\cdot\vec{x}_0}. \tag{3.4.41}$$

Examining this relationship, we see that the spectrum of fluctuations, $\tilde{f}_{\vec{k}}$, against the local background differs from the spectrum of fluctuations of the underlying field, $f_{\vec{k}}$, by a (frequency-dependent) phase factor that originates in the change of co-ordinate origin in moving from the over-all background to the local region. In addition to this phase factor, however, there is a quenching, through the factor $(1 - \sqrt{2\pi}^3 W_{\vec{k};r_*})$, of the long-wavelength modes which went into defining the local background. With regard to the frequency-dependant phase shift, since the phases of the $f_{\vec{k}}$ are uniformly distributed random variables and are not correlated with $\vec{k}$, one can drop the additional phase factor as it has no effect on the statistical properties of the fluctuations. In other words,[29]

$$\tilde{f}_{\vec{k}} = \left(1 - \sqrt{2\pi}^3 W_{\vec{k};r_*}\right) f_{\vec{k}}. \tag{3.4.42}$$

This is the key result of the present section.

Reconsidering the phase factor that was dropped in obtaining (3.4.40), some numerical tests are presented in Appendix C.6 to quantitatively demonstrate the safety of removing it. At the same time, one could instead state that what we are doing here is introducing the frequency-domain filter in (3.4.42) and that the preceding calculations merely provide the reader with an interpretation of that filter as some sort of local-structure extraction tool. In that sense, the phase factor that was dropped is not giving us an approximate spectrum, rather the spectrum is exact and its interpretation is only approximate. The reader is free to understand this operation in whichever way makes him or her most comfortable.

From the filtered spectrum in (3.4.42) we can now determine the statistical properties of $\tilde{f}_{\vec{k}}$. The mean of $\tilde{f}_{\vec{k}}$ is easily calculated and is simply

$$\left\langle \tilde{f}_{\vec{k}} \right\rangle = \left(1 - \sqrt{2\pi}^3 W_{\vec{k};r_*}\right) \langle f_{\vec{k}} \rangle = 0 \tag{3.4.43}$$

as one should expect.[30] The two-point spectral correlation function is

$$\begin{aligned} \left\langle \tilde{f}_{\vec{k}} \tilde{f}^*_{\vec{k}'} \right\rangle &= \left\langle \left(1 - \sqrt{2\pi}^3 W_{\vec{k};r_*}\right) f_{\vec{k}} \left(1 - \sqrt{2\pi}^3 W_{\vec{k}';r_*}\right) f^*_{\vec{k}'} \right\rangle \\ &= \left(1 - \sqrt{2\pi}^3 W_{\vec{k};r_*}\right) \left(1 - \sqrt{2\pi}^3 W_{\vec{k}';r_*}\right) \left\langle f_{\vec{k}} f^*_{\vec{k}'} \right\rangle \\ &= S_{\vec{k}} \left(1 - \sqrt{2\pi}^3 W_{\vec{k};r_*}\right)^2 \delta\left(\vec{k} - \vec{k}'\right) \end{aligned} \tag{3.4.44}$$

where the realness of $W_{\vec{k};r_*}$ has been used.

One last point to make before moving on is that while $\overline{f}$ is a Gaussian random variable, as was pointed out at the end of the last section we are only interested in regions that appear identical to the underlying field on average. This means that we are only interested in regions for which $\overline{f} = 0$. Within such regions, (3.4.36) tells us that the local and underlying fields are identical, i.e. $\tilde{f} = f$. Let us now make the following replacement: take $\tilde{f}$ to be the underlying field and examine the statistics of smoothed regions within it instead of within f. What we have accomplished in doing this is the replacement of the underlying field with a version of it that has had long wavelength fluctuations filtered out. This is similar to the idea of placing the field in a large box to impose a longest wavelength to its structure but in this case the spectrum remains continuous and the resulting large-scale periodicity is only approximate.

---

[29] We are really only making use of the field's homogeneity: its statistical properties are the same everywhere, so we might as well compute them at the one spot where the phase factor is equal to 1.

[30] This result holds without the constraint on the value of $\overline{f}$.

3.4.4 The Properties of $\overline{f}$, Part II

We now have the statistical properties of the random field when it is expressed with respect to its mean value smoothed over a region of radius r_*. We have called this field $\tilde{f}$. If, as discussed in the previous section, we use $\tilde{f}$ rather than f in (3.4.29) and reconsider the variance of the smoothed field we find

$$\begin{aligned}\left\langle \overline{f}^2(\vec{x}_0; r)\right\rangle &= \int W_{\vec{k};r} W_{\vec{k}';r} \left\langle \tilde{f}_{\vec{k}} \tilde{f}^*_{\vec{k}'}\right\rangle e^{i\left(\vec{k}-\vec{k}'\right)\cdot\vec{x}_0}\, d^3k\, d^3k' \\ &= \int W_{\vec{k};r} W_{\vec{k}';r} S_{\vec{k}} \left(1 - \sqrt{2\pi}^3 W_{\vec{k};r_*}\right)^2 \delta\left(\vec{k}-\vec{k}'\right) e^{i\left(\vec{k}-\vec{k}'\right)\cdot\vec{x}_0}\, d^3k\, d^3k' \\ &= \int W^2_{\vec{k};r} \left(1 - \sqrt{2\pi}^3 W_{\vec{k};r_*}\right)^2 S_{\vec{k}}\, d^3k. \end{aligned} \tag{3.4.45}$$

Again, let us use the field's isotropy and do the angular integration to get

$$\begin{aligned}\left\langle \overline{f}^2(\vec{x}_0; r)\right\rangle &= 4\pi \int W^2_{k;r} \left(1 - \sqrt{2\pi}^3 W_{k;r_*}\right)^2 S_k k^2\, dk \\ &= (2\pi)^3 \int W^2_{k;r} \left(1 - \sqrt{2\pi}^3 W_{k;r_*}\right)^2 \mathcal{P}(k) \frac{dk}{k}. \end{aligned} \tag{3.4.46}$$

Compare this last result to (3.4.32). In doing so, there are two view points one can take. On the one hand we can consider ourselves to have introduced an alternative power spectrum given by $\left(1 - \sqrt{2\pi}^3 W_{k;r_*}\right) \mathcal{P}(k)$; or, on the other hand, we can say that we have introduced a new window function whose Fourier transform is given by $W_{k;r}\left(1 - \sqrt{2\pi}^3 W_{k;r_*}\right)$. This window function has two parameters: r_*, the size of the smoothing done to establish the local effective background, and r, the size of the smoothing done to generate $\overline{f}$ with $r < r_*$. The spectral behaviour of this window function is shown in Figure 3.6 where its logarithmic amplitude is shown as a function of kr for $r_*/r = 10$ for both top-hat and Gaussian windows.

The question is now that of what r_* should be. There are several criteria by which a selection can be made. On the one hand, one could base one's choice on an analysis of the accuracy of the model for perturbations and set r_* to the length scale beyond which the model is not valid. If considering inflaton perturbations, for example, one could set r_* to the current size of the mode that left the de Sitter horizon at the earliest time for which the perturbation model is accurate.

On the other hand, one could set r_* to the length scale beyond which one believes it

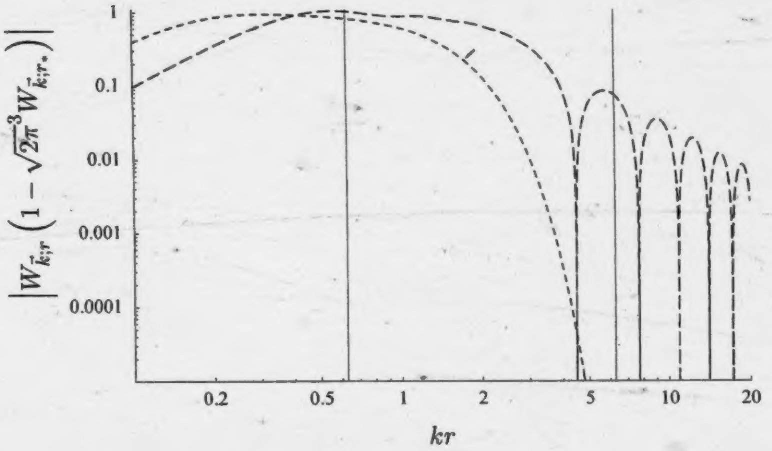


Figure 3.6: The spectral behaviour of $W_{\vec{k};r}\left(1-\sqrt{2\pi}^3 W_{\vec{k};r_*}\right)$ for $r_*/r = 10$ illustrating the long wavelength filtering. The short dashes show Gaussian windows while the long dashes show top-hat windows. The horizontal axis is kr with the reciprocal smoothing radii indicated, from left-to-right, at $k = 2\pi/r_*$ and $k = 2\pi/r$. The vertical axis is the same as that in Figure 3.4.

should no longer matter what r_* is. This second choice is based on the ergodicity of the perturbation field — the property that a single realization of the field contains sufficient information to fully characterize the ensemble — and the belief that at some large scale there will be a genuine cut-off to the fluctuations. When a field is ergodic, ensemble averages are equivalent to spatial averages over a single realization. An infinite Gaussian random field is ergodic so our perturbation field has this property. This means that one can replace the ensemble average of $\overline{f}^2(\vec{x}_0; r)$ with an average over choice of centre, $\vec{x}_0$. Now, although our perturbation field has been taken to be infinite, we are asking if it is really necessary to preform an average over the whole thing in order to determine the statistics of some quantity or if, instead, there is some size beyond which one can say an average is close enough to the true expectation value. It is possible to quantify how much an average over a limited sample size can be expected to deviate from the true average and this is called the standard deviation of the mean.

Consider some process that generates an infinite sequence of independent Gaussian-distributed samples, x_i, whose mean value is $\langle x \rangle$, and where the samples are spread about the mean with a variance of $\left\langle (x_i - \langle x \rangle)^2 \right\rangle = \sigma^2$. Now consider computing an average of n of the samples starting at randomly-chosen sample number i_0,

$$\overline{x}_{i_0} = \langle x_{i_0}, \ldots, x_{i_0+n-1} \rangle = \frac{1}{n} \sum_{i=i_0}^{i_0+n-1} x_i. \tag{3.4.47}$$

Of course, generally this average will not equal the true mean of the sequence but one can easily determine its distribution over choices of i_0. Its expectation value is clearly the mean of the samples,

$$\langle \overline{x}_{i_0} \rangle_{i_0} = \langle x \rangle \,, \tag{3.4.48}$$

while its variance is[31]

$$\left\langle \left(\overline{x}_{i_0} - \langle \overline{x}_{i_0} \rangle_{i_0} \right)^2 \right\rangle_{i_0} = \frac{1}{n}\sigma^2 . \tag{3.4.49}$$

In our case, we have a process, namely choosing $\vec{x}_0$, that generates samples of $\overline{f}^2$. Averaging this over n samples will result in a Gaussian random variable with an expectation value of $\left\langle \overline{f}^2 \right\rangle$, the true mean square for the field, and a variance of $\frac{1}{n}\left(\left\langle \overline{f}^4 \right\rangle - \left\langle \overline{f}^2 \right\rangle^2 \right)$. For Gaussian random variables, the fourth moment is related to the first and second moments by[32]

$$\begin{aligned} \left\langle x^4 \right\rangle &= 4\left(3\left\langle x^2 \right\rangle \langle x \rangle - 2\langle x \rangle^3 \right) \langle x \rangle + 3\left\langle x^2 \right\rangle^2 - 12\left\langle x^2 \right\rangle \langle x \rangle^2 + 6\langle x \rangle^4 \\ &= 3\left\langle x^2 \right\rangle^2 - 2\langle x \rangle^4 \end{aligned} \tag{3.4.50}$$

so since the mean of $\overline{f}$ is zero, the variance of the distribution for the average of n $\overline{f}^2$'s is

$$\frac{1}{n}\left(\left\langle \overline{f}^4 \right\rangle - \left\langle \overline{f}^2 \right\rangle^2 \right) = \frac{2}{n}\left\langle \overline{f}^2 \right\rangle^2 . \tag{3.4.51}$$

The standard deviation of the distribution as a fraction of the expectation value is, then, simply $\sqrt{\frac{2}{n}}$.

For example, if one wishes the standard deviation of the mean to be 1% of the mean then 20000 samples are required while if it is acceptable that the standard deviation be 10% of the mean then only 200 samples are required. In the latter case, a volume with a radius 4 times the size of the region smoothed to get $\overline{f}$ would suffice while in the former case the radius must be 17 times as large. In any case, it is clear that r_* need not be much more than one or two orders of magnitude larger than r in order to get reasonable statistics for $\overline{f}$ if, and this must be emphasized, there is truly a largest scale to the fluctuations.

The reason it is important there be a largest scale to the fluctuations becomes clear when one considers that for a $n_s = 1$ power spectrum, the true mean of $\overline{f}^2$ is infinite. In that case it is not clear what one is saying when one asserts that an average over a limited sample will not differ from the true average by more than some percentage. Below, for example, we will succeed is obtaining a finite result for $\left\langle \overline{f}^2 \right\rangle$ — a result that differs from the true, infinite, value by an infinite amount. The argument is that choosing

---

[31] It is in obtaining this result that the samples need to be independent otherwise the cross terms cannot be neglected. We have also used the ergodicity of infinite Gaussian processes to replace the average over choice of starting index, i_0, with an ensemble average.

[32] This is obtained from the last two equations in [92, equation (2.9)] and also using the fact that the third and fourth cumulants for a Gaussian distribution are 0.

r_* in the manner being described here is not intended to give results that are close to those obtained from the ideal $n_s = 1$ spectrum, which they clearly are not, but rather close to those obtained from the anticipated physical spectrum.

Returning to (3.4.46), choosing the same power-law spectrum as in (3.4.33) and taking both window functions to be Gaussian gives

$$\begin{aligned}\left\langle \bar{f}^2 \right\rangle &= (2\pi)^3\sigma^2 \int_{k=0}^{\infty} \left[\frac{1}{\sqrt{2\pi}^3}\exp\left(-\frac{1}{2}k^2r^2\right)\right]^2 \left[1-\exp\left(-\frac{1}{2}k^2{r_*}^2\right)\right]^2 k^{n_s-2}\,\mathrm{d}k \\ &= \frac{\sigma^2}{r^{n_s-1}} \int_{x=0}^{\infty} \exp\left(-x^2\right)\left[1-\exp\left(-\frac{1}{2}\frac{{r_*}^2}{r^2}x^2\right)\right]^2 x^{n_s-2}\,\mathrm{d}x. \qquad (3.4.52)\end{aligned}$$

This proves to be tricky to integrate. The results given in the tables of integrals I have examined are arrived at by expanding the $[1-\exp(\cdots)]^2$ term using the binomial theorem and then performing the entire integral term-by-term. This is fine for $n_s > 1$ where the integral of each term converges individually and, of course, this is quoted as a condition in the tables. Since the original unfiltered spectrum for $\left\langle \bar{f}^2 \right\rangle$ converged for $n_s > 1$ anyway, it does not help us to only know the result of the present integral in that regime. Our concern is how the result behaves for $n_s \leq 1$ where, it should be noted, this integral does still converge. The domain of spectral indices for which the integral converges at small k can be found by checking

$$\lim_{x\to 0_+} \left[1-\exp(-ax^2)\right]^2 x^{n_s-2} = \begin{cases} 0 & ,n_s > -2 \\ a^2 & ,n_s = -2 \\ \infty & ,n_s < -2 \end{cases},$$

which tells us the integral converges at small k for all $n_s \geq -2$.

Shown in Figure 3.7 is the result of using `Mathematica`'s numerical integration routines to evaluate the integral. This figure shows a repeat of the analysis displayed in Figure 3.5 but this time comparing the effects of long wavelength smoothing at different scales. It is clear that smoothing the long wavelength modes has had the desired effect of bringing $\left\langle \bar{f}^2 \right\rangle$ under control for $n_s \leq 1$. It is also important to note that at $n_s = 1$, going from $\frac{r_*}{r} = 10$ to $\frac{r_*}{r} = 100$ introduces a change in $\left\langle \bar{f}^2 \right\rangle$ of only about a factor of 2. It is good that the result is not overly sensitive to our choice of r_* within the range of useful values. A more detailed examination of the behaviour of $\left\langle \bar{f}^2 \right\rangle$ as a function of $\frac{r_*}{r}$ specifically for $n_s = 1$ is shown in Figure 3.8. This clearly shows the logarithmic increase of $\left\langle \bar{f}^2 \right\rangle$ with increasing $\frac{r_*}{r}$. Of course, the result never stops increasing in this case since it must diverge as $\frac{r_*}{r} \to \infty$. One has postulated, however, the existence of a length scale at which the fluctuation spectrum drops off. In reality, then, one will eventually reach this length scale at which by which point the curve shown in Figure 3.8 must have levelled off. The choice of r_* should be made, then, based on how far one feels one needs to go before things are "close enough."

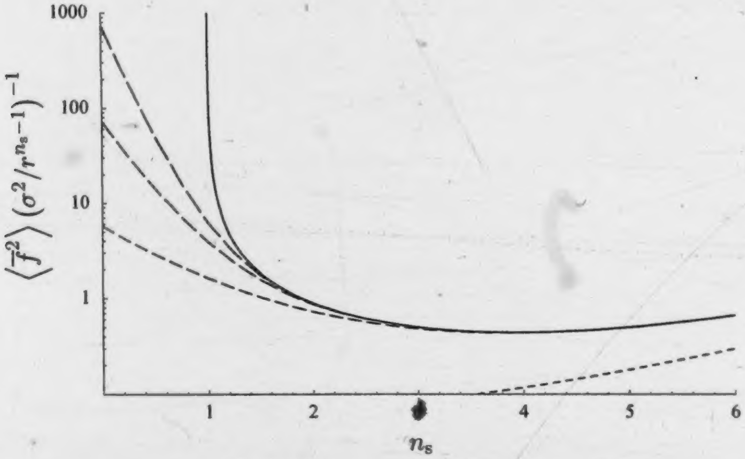


Figure 3.7: The behaviour of $\left\langle \bar{f}^2 \right\rangle \left(\sigma^2/r^{n_s-1}\right)^{-1}$ as a function of the spectral index n_s for four choices of $\frac{r_*}{r}$ using Gaussian windows. The four dashed curves in order of increasing dash length show the results for $\frac{r_*}{r} = 1, 10, 100$, and 1000 respectively. The solid curve shows the result for $\frac{r_*}{r} = \infty$ which recovers the behaviour shown in Figure 3.5 for the Gaussian window.

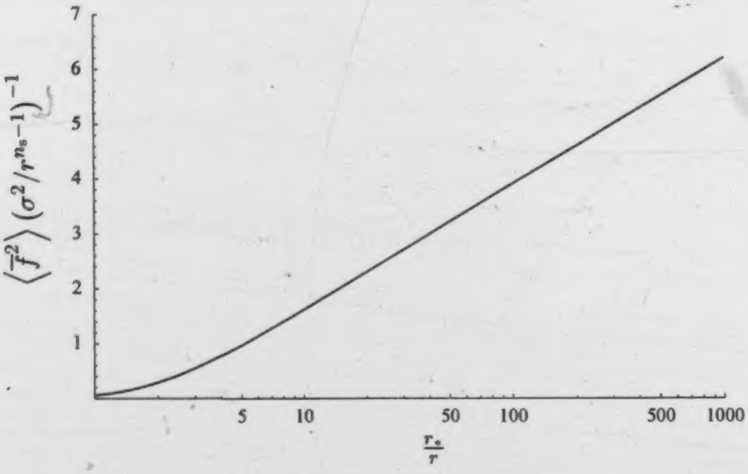


Figure 3.8: The behaviour of $\left\langle \bar{f}^2 \right\rangle \left(\sigma^2/r^{n_s-1}\right)^{-1}$ as a function of $\frac{r_*}{r}$ for a fluctuation spectrum with $n_s = 1$ illustrating the logarithmic divergence of the mean square with increasing r_*.

3.5 Matter Distribution

Taking $\rho(\vec{x})$ to describe the distribution of the material content of the universe's energy density on a spatial 3-surface, we will decompose it into a background and perturbation by writing

$$\rho(\vec{x}) = \rho_{\mathrm{b}}\,[1 + \delta|_{\vec{x}}] \tag{3.5.1}$$

where ρ_{b} is the background energy density averaged over the entire hyper-surface and δ is called the density contrast. The density contrast is position dependent but since the notation $\delta(\vec{x})$ is too easily confused with a Dirac-δ function, the position dependence of the δ's will either remain implied or be indicated via the "evaluated at" operator as above.

We can construct a smoothed version of the density contrast field, $\overline{\delta}$, and this smoothing can either be performed on the energy density itself or on the density contrast. The equivalence of these procedures is easily demonstrated. Using (3.4.5) to construct the smoothed density contrast field we find

$$\begin{aligned}
\overline{\delta}\big|_{\vec{x}_0;r} &= \int_{\vec{x}} W(\vec{x}_0 - \vec{x}; r)\, \delta|_{\vec{x}}\, \mathrm{d}^3x \\
1 + \overline{\delta}\big|_{\vec{x}_0;r} &= \int_{\vec{x}} W(\vec{x}_0 - \vec{x}; r)\, \mathrm{d}^3x + \int_{\vec{x}} W(\vec{x}_0 - \vec{x}; r)\, \delta|_{\vec{x}}\, \mathrm{d}^3x \\
\rho_{\mathrm{b}}[1 + \overline{\delta}\big|_{\vec{x}_0;r}] &= \int_{\vec{x}} W(\vec{x}_0 - \vec{x}; r)\, \rho_{\mathrm{b}}[1 + \delta|_{\vec{x}}]\, \mathrm{d}^3x \\
\rho_{\mathrm{b}}[1 + \overline{\delta}\big|_{\vec{x}_0;r}] &= \int_{\vec{x}} W(\vec{x}_0 - \vec{x}; r)\, \rho(\vec{x})\, \mathrm{d}^3x = \overline{\rho}(\vec{x}_0; r)
\end{aligned}$$

where the normalization of the window functions is used to get the second line. The last line shows that the smoothed energy density and smoothed density contrast are related to one another in exactly the manner of (3.5.1).

Within the region centred on $\vec{x}_0$, we can introduce a local density contrast, $\tilde{\delta}$, as in (3.4.7) to describe the true density contrast as a departure from the smoothed value at $\vec{x}_0$,

$$\delta|_{\vec{x}_0+\vec{x}'} = \overline{\delta}\big|_{\vec{x}_0} + \tilde{\delta}\Big|_{\vec{x}'}\,. \tag{3.5.2}$$

Substituting (3.5.2) into (3.5.1) shows that

$$\rho(\vec{x}_0 + \vec{x}') = \rho_{\mathrm{b}}\left[1 + \overline{\delta}\big|_{\vec{x}_0} + \tilde{\delta}\Big|_{\vec{x}'}\right] \tag{3.5.3}$$

Chapter 4

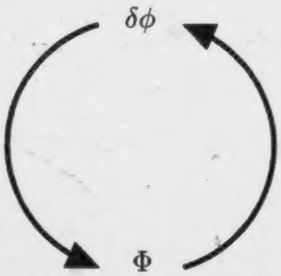
The Behaviour of Cosmological Perturbations

Returning to the analysis in Chapter 2, the field that drives inflation has, thus far, been considered to be a homogeneous field. We must now introduced perturbations to this field. In addition to describing the perturbations to the inflaton field itself, however, we will also need to describe how the geometry of space-time becomes distorted as a result of fluctuations in the material content of the universe. Not only are these two perturbations related to one another but their time evolutions are, of course, inexorably intertwined with one-another via the Einstein field equation, (2.1.1), and the field's Euler-Lagrange equation, (2.2.2).

The equations governing the behaviour of the perturbations are quite complex and, in general, cannot be solved analytically. We must approximate the solutions. The most important approximation is to assume the perturbations are small. With this approximation, the behaviour of the perturbations will only be studied to first order in the fields. We are still left with the problem of obtaining solutions for the inflaton perturbations *and* the geometry perturbations that are consistent with one another. Geometry perturbations affect the solution for the inflaton field via the form of the derivative operators in the Euler-Lagrange equation, (2.2.2). Inflaton perturbations, in turn, affect the geometry of space-time via the Einstein equation, (2.1.1).

We will approach the consistency problem in stages. To begin with, solutions for the inflaton perturbations will be obtained ignoring back-reaction on the geometry of space-time. In other words, we will solve for the behaviour of inflaton perturbations in the unperturbed Friedmann-Robertson-Walker background. We will then take the inflaton perturbations and use them as a source term in determining the perturbations to the geometry of space-time. In principle this process could be iterated, each time refining the solutions for the two perturbation fields but this will not be done here. The procedure is illustrated in the following diagram.

Use inflaton perturbation as source term and solve for geometry perturbation, Φ.



Use geometry perturbation to determine $\Box$ and solve for inflaton perturbation, $\delta\phi$.

The actual perturbation fields have not yet been defined but this is a purely schematic diagram so their precise definitions are unimportant at this stage. We will start at the bottom of the diagram with the metric perturbations assumed to be 0 and make one complete cycle; thus, we will not be left with a fully consistent solution for the perturbations, even to first order, but alas this is all we can afford.

The purpose of discussing the details of the generation and behaviour of cosmological perturbations in the present document is to explain exactly how the fluctuation spectra that will be used in analyzing the distribution of collapse times are obtained. In principle these could simply be quoted but due to the wide array of notations and gauge choices in the literature, it is once again useful to establish a common ground at this point. This is not only the case for the perturbation spectra but also for the equations of motion which will be needed for the collapse-time calculations. Finally, since the collapse calculations involve extrapolating observed spectra to length scales much larger than those used in determining them, it is important to see why it is reasonable to do this.

4.1 Quantum Inflaton Perturbations

As advertised, then, the first step is to determine the behaviour of inflaton perturbations in the unperturbed Friedmann-Robertson-Walker background. This treatment has been presented by many different authors in many different forms for various models of inflation. See, for example, [8], [31], [32], and [55, Chapter 7].

We have been treating the inflaton field classically. As far as the gross behaviour of the field is concerned, classicality is reasonable since it is assumed that the field initially takes on some tremendously large value within the region of interest — the large value corresponding to a macroscopic number of field quanta. The fluctuations, however, must be small in order for linear perturbation theory to hold so we must consider them to be a quantum system. The quantum treatment of fluctuations in the inflaton field will be accomplished by considering the field, ϕ, to be composed of two parts:

$$\phi = \phi_0 + \varphi. \tag{4.1.1}$$

The field is assumed to be dominated by a background field ϕ_0 which is treated classically. The fluctuations, φ, are assumed to be small perturbations to the classical field and are treated quantum mechanically. Substituting this expression for ϕ into (2.2.1) gives for the Lagrangian

$$L = -\frac{1}{2}g^{\mu\nu}(\phi_0 + \varphi)_{;\mu}(\phi_0 + \varphi)_{;\nu} - V_0(\phi_0 + \varphi). \tag{4.1.2}$$

The potential for the field has been relabelled V_0 to associate it with what is now the background field, ϕ_0. In order that ϕ be a solution of the equation of motion, the action must be stationary under small (linear) perturbations to it. Discarding these terms from the Lagrangian leaves

$$\begin{aligned} L &= \left[-\frac{1}{2}g^{\mu\nu}\phi_{0\,;\mu}\phi_{0\,;\nu} - V_0(\phi_0)\right] + \left[-\frac{1}{2}g^{\mu\nu}\varphi_{;\mu}\varphi_{;\nu} - V(\varphi)\right] \\ &= L_0 + L_\varphi, \end{aligned}$$

where

$$V(\varphi) = V_0(\phi_0 + \varphi) - V_0(\phi_0).$$

Also,

$$V(\varphi) \approx \frac{1}{2!}\left.\frac{\mathrm{d}^2 V_0}{\mathrm{d}\phi^2}\right|_{\phi=\phi_0}\varphi^2, \tag{4.1.3}$$

the classical potential for the field φ, has been introduced and is obtained from the lowest order contributing term (the second order term) in the Taylor expansion of the background potential.[1] The Lagrangian for the classical homogeneous background component of the field ϕ (the behaviour of which has thus been the topic of consideration so far) is

$$L_0 = -\frac{1}{2}g^{\mu\nu}\phi_{0\,;\mu}\phi_{0\,;\nu} - V_0(\phi_0), \tag{4.1.4}$$

while

$$L_\varphi = -\frac{1}{2}g^{\mu\nu}\varphi_{;\mu}\varphi_{;\nu} - V(\varphi) \tag{4.1.5}$$

is the Lagrangian for the quantum field φ describing the fluctuations. By comparison with (2.2.1) and (2.3.11), the Euler-Lagrange equation of motion for φ is

$$\ddot{\varphi} + 3H\dot{\varphi} - \frac{1}{a^2}\triangle\varphi = -\frac{\mathrm{d}V}{\mathrm{d}\varphi}. \tag{4.1.6}$$

If the potential for ϕ is assumed to be relatively flat, then $\mathrm{d}^2V_0/\,\mathrm{d}\phi^2$ from (4.1.3) is small in (4.1.6) so the right-hand-side is

$$\frac{\mathrm{d}V}{\mathrm{d}\varphi} \approx 0. \tag{4.1.7}$$

Also, since we are considering Robertson-Walker space with $k = 0$, the 3-dimensional covariant Laplacian, $\triangle$, is simply the Euclidean Laplacian, ∇^2. The result is

$$\ddot{\varphi} + 3H\dot{\varphi} - \frac{1}{a^2}\nabla^2\varphi = 0. \tag{4.1.8}$$

---

[1] Recall the linear order term was discarded from the Lagrangian.

The problem of studying the quantum behaviour of the inflaton field ϕ is thus reduced to studying the behaviour of the massless scalar field φ whose equation of motion is (4.1.8). We will assume that H is approximately constant so that space-time has approximately de Sitter geometry.

The quantization of fields in the de Sitter background is treated in detail in [16] and the applications of the results to inflation can be found in [60, Section 7.3], [75, Section 17], and [94] among others. The scalar field operator $\varphi(\vec{x}, t)$ can be decomposed into modes as follows.

$$\begin{aligned}\varphi(\vec{x}, t) &= \int \mathrm{d}^3k \left[u_k(\vec{x}, t) a_k^\dagger + u_k^*(\vec{x}, t) a_k\right] \\ &= \frac{1}{\sqrt{2\pi}^3} \int \mathrm{d}^3k \left[\psi_k(t) \mathrm{e}^{\mathrm{i}\vec{k}\cdot\vec{x}} a_k^\dagger + \psi_k^*(t) \mathrm{e}^{-\mathrm{i}\vec{k}\cdot\vec{x}} a_k\right] \qquad (4.1.9)\end{aligned}$$

where the mode functions have been separated into the form

$$u_k(\vec{x}, t) = \frac{1}{\sqrt{2\pi}^3} \psi_k(t) \mathrm{e}^{\mathrm{i}\vec{k}\cdot\vec{x}}, \qquad (4.1.10)$$

and $\vec{k}$ is the conformal momentum of the modes[2] which is related to the physical momentum of the modes, $\vec{p}$, by the Robertson-Walker scale factor[3]

$$\vec{p} = \frac{1}{a(t)} \vec{k}, \qquad (4.1.11)$$

$$\mathrm{d}^3p = \frac{1}{a^3(t)} \mathrm{d}^3k. \qquad (4.1.12)$$

Substituting (4.1.9) into (4.1.8) gives

$$\ddot{\psi}_k(t) + 3H\dot{\psi}_k(t) + H^2k^2\mathrm{e}^{-2Ht}\psi_k(t) = 0 \qquad (4.1.13)$$

as the equation of motion for the mode functions, $\psi_k(t)$.[4] This can be solved in closed form which is done in Appendix C.7. The solution is

$$\psi_k(\eta) = \frac{\sqrt{\pi}}{2} H \eta^{\frac{3}{2}} \left[c_1(k)\, \mathrm{H}_{3/2}^{(1)}(k\eta) + c_2(k)\, \mathrm{H}_{3/2}^{(2)}(k\eta)\right] \qquad (4.1.14)$$

where η is the conformal time co-ordinate, $\mathrm{H}_{3/2}^{(i)}$ are the Hankel functions of the first and second kind of order 3/2, and the solution is normalized by requiring $|c_1|^2 - |c_2|^2 = 1$.[5]

---

[2] From (4.1.9) we see that $\vec{k}$ is in radians per co-ordinate interval not radians per physical length which is why $\vec{k}$ is called "conformal."

[3] Not to be confused with the annihilation operator, a_k.

[4] Compare to [60, equation (7.3.5)] and note that Linde is using an $a(t)$ that is dimensionless.

[5] The normalization condition arises from demanding that each particle have unit probability of being found somewhere on any given space-like hyper-surface and, therefore, for the divergence of its four-current to be 0.

In Minkowski space-time, the standard solution to the equivalent of (4.1.13) is

$$\psi_k(t) = \frac{1}{\sqrt{2k}} e^{-ikt}. \tag{4.1.15}$$

One expects the Minkowski and de Sitter solutions to be identical in the high-frequency limit, $k \to \infty$, since on short distance scales de Sitter space-time looks like Minkowski space-time. This requirement can be satisfied if $c_1(k) \to 1$ and $c_2(k) \to 0$ as $k \to \infty$. See Appendix C.7 for the details. The easiest way to ensure this is to simply choose $c_1(k) = 1$ and $c_2(k) = 0$ for all k. The small-k behaviour of $c_1(k)$ and $c_2(k)$ would be determined by the initial conditions of the inflating universe and, according to [60, page 158], as long as the inflationary period is long enough the details of their behaviour are unimportant — these wavelengths get red-shifted well beyond any physically or theoretically relevant length scale. And in any case, if we end up high-pass filtering the field as in Section 3.4.3 then these long wavelength modes are certainly of no consequence.

Substituting η, c_1, c_2 and the explicit forms of the Hankel functions into $\psi_k(\eta)$ gives[6]

$$\begin{aligned}\psi_k(t) &= \frac{\sqrt{\pi}}{2} H \left(-e^{-Ht}\right)^{\frac{3}{2}} \left(-\sqrt{\frac{2}{-\pi k e^{-Ht}}}\right) \left(1 - \frac{1}{ik e^{-Ht}}\right) \exp\left(ik e^{-Ht}\right) \\ &= \frac{-iH}{k\sqrt{2k}} \left(1 - ik e^{-Ht}\right) \exp\left(ik e^{-Ht}\right).\end{aligned} \tag{4.1.16}$$

This result is exact in so far as H is a constant, (4.1.7) holds, and back-reaction on the geometry, $g_{\mu\nu}$, is ignored. As pointed out in [55, Section 7.4.4] and in [60, page 158], the important thing to notice about (4.1.16) is that as t grows, the oscillations of $\psi_k(t)$ exponentially red-shift to zero frequency and that as $t \to \infty$, $\psi_k(t)$ approaches an asymptotic value of

$$\psi_k(t) \overset{t\to\infty}{\longrightarrow} \frac{-iH}{k\sqrt{2k}}. \tag{4.1.17}$$

This can be used as a reasonable approximation for $\psi_k(t)$ once $ke^{-Ht} = k/(aH) \ll 1$ which becomes true within a few e-foldings of the mode leaving the horizon. The is the phenomenon referred to in the literature as the "freezing-in" of quantum fluctuations — the time evolution of each mode ceases shortly after expanding beyond the size of the de Sitter horizon.

The mean-square amplitude of the fluctuations of the quantum perturbations of the inflaton field, $\langle\varphi^2\rangle$, are related to ψ_k by a simple relation, namely[7]

$$\langle\varphi^2\rangle = \frac{1}{(2\pi)^3} \int |\psi_k(t)|^2 \, d^3k. \tag{4.1.18}$$

---

[6] Compare this result to [55, equation (7.81)] but note that they are working in a universe that is a box of size L; also compare with [60, equation (7.3.8)] but note the differences in Linde's notation pointed out in Appendix C.7 and his (ambiguous) assignment of dimensions that causes factors of H to move around.

[7] Compare to [60, page 158].

See Appendix C.8. By comparing (4.1.18) with the definition of the spectral density function in (3.1.14) and (3.1.15) we see that $|\psi_k|^2$ plays the role of S_k. It must be remembered, however, that we are dealing with a quantum field rather than a classical random field so, for the moment, one should only note the analogy rather than make an identification.

4.2 The Transition from Quantum to Classical Perturbations

There is not at this time a complete picture of how the apparently well-defined, classical, cosmological structure we observe around us derived itself from the quantum fluctuations it is believed to have originated in. Somewhere along the way it lost its "quantum" character and how, precisely, this happened remains a mystery. Unfortunately, the present document cannot fully escape from this subject so it will be touched on briefly.[8]

The essential element of the present-day treatments is to simply introduce a classical (in the sense that it is devoid of any special commutation properties) random field and assign to it characteristics such that it possesses the same statistical properties as the prototype quantum system. Our universe is then taken to be one realization of the classical random field. Which particular properties are chosen to be mirrored in the classical system and how exactly they are defined for the quantum system differs from treatment to treatment.

Here we will make the following construction. From above, φ is the quantum field describing the inflaton perturbations, and we will introduce $\delta\phi$ to be the classical random field mirroring the perturbations. For the quantum field, modes with distinct conformal wavenumber are eigenstates of the system so the expectation of the projection of any mode with fixed conformal wave-number onto another is 0. We will choose to mirror this property in the classical random field by making it homogeneous since this leads to the two-point correlation function of the random field's Fourier transform being a δ-function as in (3.1.10). This, in turn, leads to the conclusion that the field is Gaussian. Since the mode functions for the quantum field, $\psi_k(t)$, depend only on $k = \left|\vec{k}\right|$, the field's observables are invariant under a rotation of the field so we will also choose our random field to be isotropic. We will set the mean of our homogeneous isotropic Gaussian random field to 0,

$$\langle \delta\phi \rangle = 0, \tag{4.2.1}$$

as any non-zero mean should accounted for by the homogeneous background from which this is a perturbation. This is also consistent with the idea that the quantum field is in the ground state inside the horizon. Considering the spectral decomposition of $\delta\phi$, we will identify its spectral density function with the analogous quantity for the quantum field, so

---

[8] The details of the process by which cosmological perturbations "decohere" is a field of research unto itself. For an overview, see [42], [43], and [72, chapter 10] for an overview. More detailed information can be found in [13], [18], [19], [69] and [94]. The effect of the process on the Gaussianity of cosmic fluctuations can be found in [54].

$$S_{\vec{k}}(t) = \left\langle \left|\delta\phi_{\vec{k}}\right|^2 \right\rangle = \left|\psi_{\vec{k}}(t)\right|^2$$

and from (4.1.16),

$$S_{\vec{k}}(t) = \frac{H^2}{2k^3}\left(1 + k^2 \mathrm{e}^{-2Ht}\right). \tag{4.2.2}$$

As explained in Chapter 3, the spectral density function is all that is required to fully characterize an homogeneous isotropic Gaussian random field with a mean of zero. Using the relationships found in Section 3.1, all statistical properties of the field can be derived from $S_{\vec{k}}$.

Looking at this last result reveals something that has not been discussed yet but really should be commented on at this point. It was claimed, above, that the quantum perturbations to the inflaton field start out in the ground state and essentially remain in the ground state until they expand beyond the de Sitter horizon. We see here that this really isn't the case: at $t = 0$ there is a non-zero contribution to the field's mean square from all modes so none of them are in the ground state. What is going on here? Well this is an artifact of one of the approximations made to solve the equation of motion for the quantum field. We assumed that H was constant and implicit in that is the assumption of a static space-time in the sense that time translation leaves geometry unchanged as long as the Robertson-Walker scale factor is redefined appropriately. Our system, therefore, has been set up to have no real beginning — the system isn't started anywhere — and to exhibit no real time evolution. We can check the time translation invariance of our perturbations by replacing t with $t + \Delta t$ in (4.2.2) to find that (remembering that the spectral density belongs under an integral)

$$\begin{aligned} S_{\vec{k}}(t+\Delta t)\,\mathrm{d}^3k &= \frac{H^2}{2k^3}\left[1 + \left(k^2\mathrm{e}^{-2H\Delta t}\right)\mathrm{e}^{-2Ht}\right]\mathrm{d}^3k \\ &= \frac{H^2}{2\left(k\mathrm{e}^{-H\Delta t}\right)^3}\left[1 + \left(k\mathrm{e}^{-H\Delta t}\right)^2\mathrm{e}^{-2Ht}\right]\mathrm{d}^3\left(k\mathrm{e}^{-H\Delta t}\right) \\ &= S_{\vec{k}\mathrm{e}^{-H\Delta t}}(t)\,\mathrm{d}^3\left(k\mathrm{e}^{-H\Delta t}\right) \end{aligned}$$

whence we see that our perturbations really are invariant under time translation as long as the spatial co-ordinates are appropriately rescaled.

The modes *are* asymptotically in the ground state as $k \to \infty$, deep inside the de Sitter horizon, which is consistent with the claim that at some time in the arbitrarily distant past every mode, whether currently inside or outside the horizon, was in the ground state. In the end, we aren't really interested in the details of a mode's behaviour inside the horizon but rather the amplitude it asymptotes after it has left the horizon. Within a few e-foldings of a mode with wave-number k leaving the horizon, the spectral density for that mode is approximately

$$S_{\vec{k}}(t > t_*) = \frac{H^2}{2k^3} \tag{4.2.3}$$

where t_* gives the time corresponding to several e-foldings after the mode in question has expanded beyond the horizon. This notation follows that of Liddle and Lyth in [55]. Of course, the actual value of t_* is k-dependant but for brevity this is left implied.

Coming back to our approximation of H begin constant, we know that this is not the case during the inflation epoch. See, for example, the plots in Figure 2.1. In arriving at the result for $\psi_k(t)$ in Section 4.1, it was acceptable to approximate H as a constant since in analyzing the dynamics of any one mode the time scale involved was the duration of the mode's transition from a small-scale mode in the vacuum state to a frozen-in super-horizon scale mode. Over such a period of time H does not change by much and the approximation is good.[9] From one mode's departure from the horizon to the next mode's, however, H may change appreciably and we now need to take that into consideration when specifying the spectral density function across a wide range of wavelengths. At this point, this is easily taken care of by restating (4.2.3) as

$$S_{\vec{k}} = \left.\frac{H^2(\phi)}{2k^3}\right|_{t\sim t_*} . \tag{4.2.4}$$

So the ϕ-dependence of H is indicated along with the specification that it should be evaluated at a time that is a few e-folds after the mode in question expands beyond the horizon. Since we have assumed that H doesn't change much over a few e-folds, one can also simply evaluate H right at the time the mode leaves the horizon. In other words,

$$S_{\vec{k}} = \left.\frac{H^2(\phi)}{2k^3}\right|_{k\sim aH} . \tag{4.2.5}$$

From the definition of the power spectrum in (3.2.3) and using (4.2.5) we see that, for the inflaton perturbations that have expanded beyond the horizon,[10]

$$\mathcal{P}_{\delta\phi}(k) = \left.\frac{H^2(\phi)}{4\pi^2}\right|_{k\sim aH} . \tag{4.2.6}$$

4.2.1 The "Curvature Perturbation"

It will be found to be convenient to deal not with the (classical) inflaton perturbations directly but rather with a different field that can be defined in terms of them. Let's define a new random field, $\mathcal{R}$, such that its Fourier transform is given by

$$\mathcal{R}_k \equiv - \left[\frac{H(\phi)}{\dot{\phi}}\delta\phi_k\right]_{k\sim aH} . \tag{4.2.7}$$

This is equivalent to the "primordial curvature perturbation" described by Liddle and Lyth in [55, Section 4.2.3]. From the definition of $\mathcal{R}$ in (4.2.7) and using (3.4.19), we see that its power spectrum is given by[11]

---

[9] The quality of this approximation can be seen in Figure 2.1. For example, as a mode grows in size from 1/1000th the horizon to 1000× the horizon, it expands by $\sim$ 14 e-folds. When $H \approx 1$, 14 e-fold times is 14 $t_{\rm Pl}$ while the small ticks on the horizontal axis of plot (b) are intervals of 2×10^9 $t_{\rm Pl}$.

[10] Agrees with [55, equation (7.87)].

$$\mathcal{P}_{\mathcal{R}}(k) = \left[\left(\frac{H(\phi)}{\dot{\phi}} \right)^2 \mathcal{P}_{\delta\phi}(k) \right]_{k \sim aH} . \tag{4.2.8}$$

Using the power spectrum for $\delta\phi$ in (4.2.6) gives[12]

$$\mathcal{P}_{\mathcal{R}}(k) = \left[\frac{H^2(\phi)}{2\pi\dot{\phi}} \right]^2_{k \sim aH} . \tag{4.2.9}$$

Recalling that $k \sim aH$ is referring to times during the inflation epoch, we can use (2.3.13) for $H(\phi)$ and (2.3.14) for $\dot{\phi}$ to write the power spectrum for $\mathcal{R}$ in terms of the inflaton potential and its derivatives,[13]

$$\mathcal{P}_{\mathcal{R}}(k) = \frac{128\pi}{3} \left[\frac{V^3(\phi)}{V'^2(\phi)} \right]_{k \sim aH} . \tag{4.2.10}$$

4.3 Geometry Perturbations — Description

The next task is to investigate the cosmological consequences of perturbations in the inflaton field. We first need to describe how it is we will quantify perturbations to the geometry of space-time and this is the subject of the current section. In the next section we will make the connection between inflaton perturbations and geometry perturbations.

There is a bewildering array of approximation techniques used to study perturbations away from the perfectly homogeneous and isotropic Friedman-Robertson-Walker geometry of the cosmos. Each technique has been developed according to the nature of the particular problem being investigated, the method of analysis being used, and also the aesthetic tastes of the researcher in question. The presentation below will be a hybrid of the analyses given in [5], [6], [60, Section 7.5], [68], [69] and [74, Section 80].

Our procedure begins with a 3+1 foliation of space-time into space and time — an ordered sequence of 3-geometries labelled by a time co-ordinate. This procedure can be found in [65, Sections 21.4 and 21.5] where we see that we must write the metric in the form [65, equation (21.42)]

$$g_{\mu\nu} = \begin{bmatrix} {}^{(4)}g_{00} & {}^{(4)}g_{0j} \\ {}^{(4)}g_{i0} & {}^{(4)}g_{ij} \end{bmatrix} = \begin{bmatrix} (N_k N^k - N^2) & N_j \\ N_i & {}^{(3)}g_{ij} \end{bmatrix} \tag{4.3.1}$$

The 3-metric, ${}^{(3)}g_{ij}$, describes the geometry within each spatial surface. The shift vector, N_i, describes the change in spatial co-ordinates experienced by a normal world line in moving from one surface to the next[14]

---

[11] Agrees with [55, equation (4.62)].

[12] Agrees with [55, equation (7.103)].

[13] Agrees with [55, equation (7.104)] noting that they are using units in which $m_{\mathrm{Pl}} \neq 1$.

$$\mathrm{d}x^i = -N^i(t,\vec{x})\,\mathrm{d}t. \tag{4.3.2}$$

The components of N_i are raised and lowered with ${}^{(3)}g_{ij}$. The lapse function, $N(t,\vec{x})$, describes the interval of proper time that elapses along a world line normal to the surfaces when moving from t to $t+\mathrm{d}t$

$$\mathrm{d}\tau = N(t,\vec{x})\,\mathrm{d}t. \tag{4.3.3}$$

The components of the inverse metric, ${}^{(4)}g^{\mu\nu}$, are[15]

$$g^{\mu\nu} = \begin{bmatrix} {}^{(4)}g^{00} & {}^{(4)}g^{0j} \\ {}^{(4)}g^{i0} & {}^{(4)}g^{ij} \end{bmatrix} = \begin{bmatrix} -N^{-2} & N^j N^{-2} \\ N^i N^{-2} & {}^{(3)}g^{ij} - N^i N^j N^{-2} \end{bmatrix} \tag{4.3.4}$$

which is easily verified by checking that $g^{\mu\alpha}g_{\alpha\nu} = \delta^{\mu}{}_{\nu}$.

As an example, for the 3+1 split of the unperturbed $k=0$ Robertson-Walker geometry in (2.1.3) and (2.1.4), the lapse function is

$$N = 1, \tag{4.3.5}$$

the shift vector is

$$N_i = 0, \tag{4.3.6}$$

and ${}^{(3)}g_{ij}$ is diagonal and given by

$$\begin{aligned} {}^{(3)}g_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j &= a^2(t)\left[\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2\right] && (4.3.7) \\ &= a^2(t)\left[\mathrm{d}r^2 + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right]. && (4.3.8) \end{aligned}$$

Perturbations to homogeneous isotropic backgrounds can be divided into three disjoint classes.[16]

- "Tensor" perturbations which are constructed only from traceless divergenceless 3-tensors. These represent gravitational waves.
- "Vector" perturbations which are constructed from divergenceless 3-vectors and their covariant derivatives. These represent vortical motion of the cosmological fluid.
- "Scalar" perturbations which are constructed from scalars and their covariant derivatives. These represent density perturbations.

Here, we will only consider scalar perturbations and we will denote the unperturbed metric and its components by [illegible]$_{\mu\nu}$, N_0, N_{0i}, and ${}^{(3)}g_{0ij}$. The most general scalar perturbation can be parameterized in terms of four functions of t and $\vec{x}$. In the 3+1 decomposition of the geometry described above, these quantities are the perturbation to the lapse function, $\alpha(t, \vec{x})$,

---

[14] This is the definition used by Misner, Thorne, and Wheeler and is opposite in sign to that used by Bardeen in [6].

[15] See [65, equation (21.44)] and also compare to [6, equation (4)] noting the difference introduced by the sign choice made in (4.3.2).

[16] See [6] and [69].

$$N = N_0 \left(1 + \alpha\right), \tag{4.3.9}$$

the perturbation to the shift vector, $\beta(t, \vec{x})$,

$$N_i = N_{0i} + a^2 \beta_{|i}, \tag{4.3.10}$$

and the perturbations to the spatial metric tensor, $\Phi(t, \vec{x})$ and $\epsilon(t, \vec{x})$,

$${}^{(3)}g_{ij} = (1 - 2\Phi)\, {}^{(3)}g_{0ij} + 2a^2 \epsilon_{|ij} \tag{4.3.11}$$

where $X_{|i}$ is the covariant derivative of X in the three-geometry with respect to the co-ordinate i.[17] Substituting these into the metric in (4.3.1) gives, to first order,

$$\begin{aligned} g_{\mu\nu} &= \begin{bmatrix} \left(N_{0k} + a^2\beta_{|k}\right)\left({N_0}^k + a^2\beta^{|k}\right) - N_0^2\left(1 + 2\alpha\right) & N_{0j} + a^2\beta_{|j} \\ N_{0i} + a^2\beta_{|i} & (1 - 2\Phi)\, {}^{(3)}g_{0ij} + 2a^2\epsilon_{|ij} \end{bmatrix} \\ &= \begin{bmatrix} N_{0k}{N_0}^k - N_0^2 & N_{0j} \\ N_{0i} & {}^{(3)}g_{0ij} \end{bmatrix} + \begin{bmatrix} 2\left(a^2 N_{0k}\beta^{|k} - N_0^2\alpha\right) & a^2\beta_{|j} \\ a^2\beta_{|i} & -2\Phi\, {}^{(3)}g_{0ij} + 2a^2\epsilon_{|ij} \end{bmatrix} \\ &= g_{0\mu\nu} + \delta g_{\mu\nu} \end{aligned}$$

where

$$\delta g_{\mu\nu} = \begin{bmatrix} 2\left(a^2 N_{0k}\beta^{|k} - N_0^2\alpha\right) & a^2\beta_{|j} \\ a^2\beta_{|i} & -2\Phi\, {}^{(3)}g_{0ij} + 2a^2\epsilon_{|ij} \end{bmatrix}. \tag{4.3.12}$$

This expression for the perturbation to the metric tensor suggests the introduction of an alternate, fifth, scalar to replace α in directly describing the perturbation to g_{00},

$$\Psi(t, \vec{x}) = N_0^2\alpha - a^2 N_{0k}\beta^{|k}, \tag{4.3.13}$$

so

$$\delta g_{\mu\nu} = \begin{bmatrix} -2\Psi & a^2\beta_{|j} \\ a^2\beta_{|i} & -2\left(\Phi\, {}^{(3)}g_{0ij} - a^2\epsilon_{|ij}\right) \end{bmatrix}. \tag{4.3.14}$$

In this form, the notation used in this document is easily compared to that of Mukhanov et al. in [69]. Comparing to [69, equation (2.9)] and noting that they are using the $(-++)$ sign convention,[18] the correspondences listed in Table 4.1 are observed.

For a given geometry, its description as a perturbation away from a background geometry in the manner described above is not unique. One can imagine that for any given, physical, space-time geometry there is more than one way to split it into a homogeneous background plus perturbation; the result being that two distinct combinations of

---

[17] Compare to [6, equations (28)–(30)] noting the difference in the choice of the sign of Φ. This has been made in order to bring the final result into agreement with the notation of Misner, Thorne and Wheeler. Although it is not apparent, the sign of β is also opposite that of [6, equation (29)] due to the different definition of the shift vector used here.

[18] See [69, Appendix A] and Appendix A of this document for a description of this sign convention.

This Document	Bardeen [6]	Mukhanov et al. [69]
α	α	
β	$-\beta$	B
Φ	$-\phi$	ψ
ϵ	γ	E
Ψ		ϕ

Table 4.1: Relationships between the perturbation variables used in this document and those of [6] and [69].

background metric and perturbation scalars may, in fact, represent the same geometry. Indeed, a "perturbation" may accomplish nothing more than producing a new representation of the same geometry and in that sense not be a perturbation at all. This "gauge freedom" can be eliminated by placing restrictions on one or more of the perturbation scalars — choosing a gauge — thereby ensuring that any perturbation is genuine. In practise, not all choices of gauge actually succeed in fully eliminating degeneracy in the description of space-time.

It can be shown that two of the four perturbation scalars can be chosen arbitrarily.[19] There are several standard choices of gauge, two of which are described below.

- Synchronous gauge: $\beta = 0$, $\Psi = 0$. This is an example of a gauge which does not fully eliminate co-ordinate ambiguity.
- Zero shear, or longitudinal, or conformal-Newtonian gauge: $\beta = 0$, $\epsilon = 0$. This gauge fully removes all co-ordinate ambiguity.

Two other standard choices of gauge are the uniform expansion gauge in which the perturbation to the trace of the extrinsic curvature of constant t hyper-surfaces is set to 0 and the co-moving gauge in which the perturbation to the momentum density of the matter is set to 0. Neither of these two gauges is easily described using the quantities defined in this section. See [6] for a more detailed description.

Having expressed a perturbation in terms of the four perturbation scalars, one finds that performing a co-ordinate transformation will alter the functional form of the perturbation. This is true not just for scalar but also vector perturbations. Restricting oneself to co-ordinate transformations that do not alter the form of any vector perturbations (thus preserving their absence if there are none to start with), it is possible to find combinations of the perturbation scalars defined above that are invariant under such co-ordinate transformations. Two such "gauge-invariant variables" are[20]

---

[19] See [69, Section 3].

$$\Psi^{(\mathrm{gi})} = \Psi + \frac{1}{a}\left[\left(\beta - \epsilon'\right)a\right]' \tag{4.3.15}$$

$$\Phi^{(\mathrm{gi})} = \Phi - \frac{a'}{a}\left(\beta - \epsilon'\right) \tag{4.3.16}$$

where $'$ indicates differentiation with respect to conformal time, η. One should notice that in the zero-shear gauge, Ψ and Φ are exactly equal to these gauge-invariant variables. This has the important consequence that any relationships obeyed by Ψ and Φ in the zero-shear gauge must be obeyed by $\Psi^{(\mathrm{gi})}$ and $\Phi^{(\mathrm{gi})}$ independent of any co-ordinate transformation.

We will choose to use the zero shear gauge which reduces the perturbed line element to

$$g_{\mu\nu}\,\mathrm{d}x^{\mu}\,\mathrm{d}x^{\nu} = -\left(1 + 2\Psi\right)\,\mathrm{d}t^2 + a^2\left(1 - 2\Phi\right)\delta_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j. \tag{4.3.17}$$

From this line element, one can compute the Einstein tensor for this geometry to first order in the perturbations. Introducing first-order perturbations to the stress-energy tensor in a manner similar to the way in which they were introduced to the metric tensor and then forming Einstein's equation gives a set of coupled differential equations for the perturbations. If one restricts oneself to the case of a perfect fluid for which the anisotropic stress, T_{ij}, $i \neq j$, vanishes, then the off-diagonal spatial parts of Einstein's equation lead to the constraint[21]

$$\left(\Psi - \Phi\right)_{|ij} = 0, \text{ for } i \neq j. \tag{4.3.18}$$

This implies that $\Psi = \Phi$ which is seen as follows.[22] From (4.3.18) all mixed derivatives of $D = \Psi - \Phi$ must vanish which implies that D is of the form $D = \sum f_i(x_i)$. Since we are working in the zero-shear gauge, all of this must also hold for $\Psi^{(\mathrm{gi})}$ and $\Phi^{(\mathrm{gi})}$ and therefore be invariant under co-ordinate transformations. This functional form for D can only be preserved under co-ordinate transformations, however, if the f_i are linear functions. Now, the spatial average of the perturbations must vanish and the only linear functions with this property are $f_i(x_i) = 0$, therefore $D = \Psi - \Phi = 0$.

Applying the constraint $\Psi = \Phi$ to (4.3.17), the linearly-perturbed line element becomes

$$g_{\mu\nu}\,\mathrm{d}x^{\mu}\,\mathrm{d}x^{\nu} = -\left(1 + 2\Phi\right)\,\mathrm{d}t^2 + a^2\left(1 - 2\Phi\right)\delta_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j \tag{4.3.19}$$

or, in spherical co-ordinates,

$$g_{\mu\nu}\,\mathrm{d}x^{\mu}\,\mathrm{d}x^{\nu} = -\left(1 + 2\Phi\right)\,\mathrm{d}t^2 + \left(1 - 2\Phi\right)a^2(t)\left[\mathrm{d}r^2 + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right]. \tag{4.3.20}$$

This line element encodes the most general linear, anisotropic stress-free, scalar perturbations to an homogeneous isotropic FRW background. As such, the line element is suitable for describing the behaviour of any linearly-perturbed perfect fluid including the inflaton field during the inflation epoch and dust during the post-inflation epoch. Compare this result to [55, equation (14.137)], [60, equation (7.5.8)], [65, equation (18.15c)], [69, equation (3.21)].

---

[20] See [69, equation (3.13)], recalling the notation differences between Mukhanov et al. and this document.

[21] See [69, equation (5.16)].

[22] This explanation can also be found in [69].

4.4 Geometry Perturbations — Behaviour

We now need to use our classical inflaton perturbations, $\delta\phi$, as the source term on the right-hand side of Einstein's equation and solve for the metric perturbations, Φ. The program is the following. Using the perturbed FRW line element we obtained in (4.3.19), we first obtain the Einstein tensor for this geometry to linear order in the metric perturbations. Then, we expand the stress-energy tensor for the inflaton field in (2.2.3) to linear order in perturbations both to the inflaton and to the metric thus completing Einstein's field equation to linear order in the perturbations. Finally, we solve the resulting set of coupled differential equations using the inflaton perturbation spectrum from (4.2.6), or (4.2.10), as given.

In solving the system of differential equations, we will use the equations of motion for the FRW scale factor in a flat universe, i.e. (2.1.6a) and (2.1.6b) with $k = 0$, and the equation of motion for the unperturbed inflaton field, (2.2.2) with spatial derivatives eliminated, to form a background solution. These will be imposed to eliminate the homogeneous terms from the equations of motion for the metric perturbations. The calculations are done in Appendix C.9.1 and the result is the system of equations,[23]

$$\ddot{\Phi} + \left(\frac{\dot{a}}{a} - 2\frac{\ddot{\phi}}{\dot{\phi}}\right)\dot{\Phi} - \frac{1}{a^2}\nabla^2\Phi + 2\left(\frac{\ddot{a}}{a} - \left(\frac{\dot{a}}{a}\right)^2 - \frac{\dot{a}}{a}\frac{\ddot{\phi}}{\dot{\phi}}\right)\Phi = 0 \tag{4.4.1a}$$

$$\frac{1}{a}\dot{(a\Phi)}_{,i} - 4\pi(\dot{\phi}\delta\phi)_{,i} = 0 \tag{4.4.1b}$$

where $a(t)$ and $\phi(t)$ are the solutions of their respective unperturbed equations of motion, a dot is differentiation with respect to t, and $\delta\phi$ is the (classical) inflaton perturbation we have previously obtained.

It is possible to transform (4.4.1a) into the form[24]

$$u'' - \nabla^2 u - \frac{[a'/(a^2\phi')]''}{[a'/(a^2\phi')]}u = 0 \tag{4.4.2}$$

where $u = \frac{a}{\phi'}\Phi$ and a prime indicates differentiation with respect to conformal time, η. See Appendix C.9.1 for the derivation. As discussed by Mukhanov in [68] and by Linde in [60, Section 7.5], it is possible to solve this equation in the long and short wavelength limits. One considers a plane wave solution so that

---

[23] Compare these to the results in [68].

[24] Compare this to the result in [68].

$$X = \frac{1}{\sqrt{2\pi}^3} \int X_{\vec{k}} e^{i\vec{k}\cdot\vec{x}} \, d^3k \tag{4.4.3}$$

where X represents each of Φ, $\delta\phi$, and u and where $\vec{k}$ is the conformal wave-number.[25] Upon applying the mode decomposition in (4.4.3) to (4.4.2) and taking the long wavelength limit in which

$$k^2 \ll \frac{(a'/a^2/\phi')''}{(a'/a^2/\phi')} \tag{4.4.4}$$

one obtains

$$u_k'' - \frac{(a'/a^2/\phi')''}{(a'/a^2/\phi')} u_k = 0. \tag{4.4.5}$$

Since this is a linear second order differential equation, to form the complete solution we need two linearly independent solutions. Clearly one solution is

$$u_{1k} = \frac{a'}{a^2\phi'} \tag{4.4.6}$$

and we can generate the second solution from this one by doing

$$u_{2k} = u_{1k} \int^{\eta} \frac{1}{u_{1k}^2} \, d\eta' = 4\pi \left(\frac{1}{\dot{\phi}} - \frac{\dot{a}}{a^2\dot{\phi}} \int^t a \, dt' \right). \tag{4.4.7}$$

See Appendix C.10 for the derivation. Writing it in terms of Φ, the complete solution is[26]

$$\Phi_k = A_k \left(1 - \frac{\dot{a}}{a^2} \int^t a \, dt' \right) + B_k \frac{\dot{a}}{a^2}. \tag{4.4.8}$$

If $a(t) \propto e^{Ht}$, as in the inflation epoch, then

$$\Phi_k = A_k \left(1 - \frac{H}{a} \frac{a}{H} \right) + B_k \frac{H}{a} = B_k \frac{H}{a} \tag{4.4.9}$$

while if $a(t) \propto t^n$, as in the post-inflation epoch, then

$$\begin{aligned} \Phi_k &= A_k \left(1 - \frac{nt^{n-1}}{t^{2n}} \frac{1}{n+1} t^{n+1} \right) + B_k \frac{nt^{n-1}}{t^{2n}} \\ &= A_k \frac{1}{n+1} + B_k \frac{n}{t^{n+1}}. \end{aligned} \tag{4.4.10}$$

---

[25] Once again, we see the inner product of $\vec{k}$ with the co-ordinate interval $\vec{x}$ rather than with the physical interval $a\vec{x}$ appearing in the phase factor.

[26] Compare this to [68, equation (13)].

For example, for a hot ultra-relativistic gas (i.e. $w = \frac{1}{3}$ in (2.1.8)), we have from (2.1.13) that $a(t) \propto t^{1/2}$ so

$$\Phi_k = \frac{2}{3}A_k + \frac{1}{2}B_k t^{-\frac{3}{2}} \tag{4.4.11}$$

while for cold pressureless dust ($w = 0$), we have $a(t) \propto t^{2/3}$ so

$$\Phi_k = \frac{3}{5}A_k + \frac{2}{3}B_k t^{-\frac{5}{3}}. \tag{4.4.12}$$

Let's consider the following, generic scenario. Assume the universe first goes through a period of inflation and that following this the universe "reheats" and becomes filled with either hot or cold gas. Then, to within a factor of H, B_k is the spectrum for the metric perturbations at the start of the inflation epoch. In all three forms for the time dependence of $a(t)$ illustrated above, the time dependence of Φ_k is such that for sufficiently large t the B_k contribution to Φ_k becomes negligible. In all three cases, $a(t)$ grows monotonically so we have the interpretation that the expansion of the universe is diluting the primordial fluctuations. During the inflation epoch, (4.4.9), this process occurs at an exponential rate. Assuming that, as long as $a(t)$ grows monotonically, the decay of the B_k contribution is a generic property of the time dependence of Φ_k, then let us now require inflation to last sufficiently long, i.e. $Ht \gg 1$, so as to allow us to ignore that contribution completely regardless of the actual behaviour of $a(t)$. In that case, the general form for Φ_k can be approximated by the A_k term of (4.4.8) alone i.e.

$$\Phi_k = A_k \left(1 - \frac{\dot{a}}{a^2}\int_0^t a\,\mathrm{d}t'\right) \tag{4.4.13}$$

for all $a(t)$ and for all t. Compare to [60, equation (7.5.15)] and [68, equation (13)]. Converting (4.4.1b) to frequency space as in (C.9.9) then using (4.4.13) and (C.9.6a) to solve for A_k gives us

$$\begin{aligned}
\frac{1}{a}\frac{\mathrm{d}}{\mathrm{d}t}\left(a\Phi_k\right) &= 4\pi\dot{\phi}\delta\phi_k \\
A_k\frac{1}{a}\frac{\mathrm{d}}{\mathrm{d}t}\left(a - \frac{\dot{a}}{a}\int_0^t a\,\mathrm{d}t'\right) &= 4\pi\dot{\phi}\delta\phi_k \\
A_k\frac{1}{a}\left(\dot{a} - \left(\frac{\dot{a}}{a}\right)^{\cdot}\int_0^t a\,\mathrm{d}t' - \frac{\dot{a}}{a}a\right) &= 4\pi\dot{\phi}\delta\phi_k \\
A_k\frac{1}{a}4\pi\dot{\phi}^2\int_0^t a\,\mathrm{d}t' &= 4\pi\dot{\phi}\delta\phi_k
\end{aligned}$$

or[27]

$$A_k = \frac{a}{\int_0^t a\,\mathrm{d}t'}\frac{\delta\phi_k}{\dot{\phi}}. \tag{4.4.14}$$

The quantities A_k are constants so we can evaluate them for any time t with the caveat that since we are dealing with a long wavelength approximation we must restrict ourselves to times when each mode does actually have a long wavelength. Because of the form we obtained for the inflaton perturbation's power spectrum in (4.2.6), it is best to choose to evaluate A_k when the wavelength of the mode in question is on the order of the horizon size, i.e. when $k/a \sim H$; this is, of course, during the inflation epoch. Now[28]

---

[27] Compare to [60, equation (7.5.16)] and [68, equation (14)].

$$\frac{a}{\int_0^t a\,\mathrm{d}t'} = H(t)\left[1 + \frac{\dot{H}}{H^2}\left(1 + O(\frac{\dot{H}}{H^2}, \frac{\ddot{H}}{H^3})\right)\right] \tag{4.4.15}$$

but during the inflation epoch $\dot{H} \ll H^2$ and $\ddot{H} \ll H^3$, so

$$A_k \approx \left.\frac{H(\phi)}{\dot{\phi}}\delta\phi_k\right|_{k\sim aH}. \tag{4.4.16}$$

Again, to be clear: the A_k are *constants*. When, exactly, they are evaluated is irrelevant. Here, we have chosen to evaluate them at particular times during the inflation epoch purely for convenience. Notice that $-A_k$ is exactly the "curvature perturbation" defined in (4.2.7). This, finally, shows the advantage of using $\mathcal{R}_k = -A_k$ to describe the inflaton perturbations rather than, for example, $\delta\phi_k$: $\mathcal{R}_k$ is time-independent. One consequence of this, and even more justification for its use, is that in principle it can continue to be well defined even after the inflaton field has decayed away.

Substituting $A_k = -\mathcal{R}_k$ from (4.4.16) into (4.4.13) gives the relationship between long wavelength inhomogeneities in the inflaton field and the corresponding metric perturbations, namely

$$\Phi_k = -\left(1 - \frac{\dot{a}}{a^2}\int_0^t a\,\mathrm{d}t\right)\mathcal{R}_k. \tag{4.4.17}$$

This relationship shows one reason for calling $\mathcal{R}$ the "curvature perturbation," and that is that it is the time-independent part of the perturbation to the Robertson-Walker line element. Using (3.4.19),

$$\mathcal{P}_\Phi(k) = \left(1 - \frac{\dot{a}}{a^2}\int_0^t a\,\mathrm{d}t\right)^2 \mathcal{P}_\mathcal{R}(k). \tag{4.4.18}$$

In the post inflation epoch, $a(t) \propto t^n$ and recalling (4.4.10) for this case gives

$$\mathcal{P}_\Phi(k) = \frac{1}{(n+1)^2}\mathcal{P}_\mathcal{R}(k). \tag{4.4.19}$$

Using (2.1.13) to express n in terms of w, the ratio of pressure to energy density,

$$\mathcal{P}_\Phi(k) = \frac{9}{25}\left(\frac{1+w}{1+\frac{3}{5}w}\right)^2 \mathcal{P}_\mathcal{R}(k). \tag{4.4.20}$$

---

[28] See [60, page 178].

From (C.9.46) in Appendix C.9.2, we see that for modes inside the Hubble horizon $(k/(aH) \gg 1)$ of a flat, pressureless, universe, the energy density contrast is related to the metric perturbation by

$$\delta_k \approx -\frac{2}{3}\left(\frac{k}{aH}\right)^2 \Phi_k. \tag{4.4.21}$$

Using (3.4.19) gives

$$\mathcal{P}_\delta(k) = \frac{4}{9}\left(\frac{k}{aH}\right)^4 \mathcal{P}_\Phi(k), \tag{4.4.22}$$

and using (4.4.20) with $w = 0$ (cold dust) we get[29]

$$\mathcal{P}_\delta(k) = \left(\frac{k}{aH}\right)^4 \frac{4}{25}\mathcal{P}_\mathcal{R}(k). \tag{4.4.23}$$

Using (4.2.10) for $\mathcal{P}_\mathcal{R}(k)$ gives[30]

$$\mathcal{P}_\delta(k) = \left(\frac{k}{aH}\right)^4 \frac{2\pi}{3}\left(\frac{16}{5}\right)^2 \left[\frac{V^3(\phi)}{V'^2(\phi)}\right]_{k\sim aH}. \tag{4.4.24}$$

4.4.1 Refinement of Inflaton Perturbations

Although this will not be done here, for the curious the next step in the iterative refinement of the cosmological perturbations is to use the metric perturbations to obtain more accurate results for the inflaton perturbations. This is done by now writing the Euler-Lagrange equation of motion for the inflaton field, (2.2.2), to linear order in perturbations both to the field and to the metric. The result, derived in Appendix C.9.1, is[31]

$$\delta\ddot{\phi} + 3\frac{\dot{a}}{a}\delta\dot{\phi} - \frac{1}{a^2}\nabla^2\delta\phi + \frac{\mathrm{d}^2V}{\mathrm{d}\phi^2}\delta\phi + 2\frac{\mathrm{d}V}{\mathrm{d}\phi}\Phi - 4\dot{\Phi}\dot{\phi} = 0 \tag{4.4.25}$$

where $a(t)$ and $\phi(t)$ are the solutions of their respective unperturbed equations of motion. Notice the two additional terms linear in the metric perturbations that are present in this equation but not in (4.1.6). It is the absence of these terms from the equation that was solved to obtain the inflaton perturbations that prevents us from claiming our solutions are consistent with one-another to first order.

This equation, (4.4.25), together with the linearly perturbed Einstein equation in (4.4.1) constitute a single set of coupled differential equations and are the complete form of the equations accurate to linear order in the perturbations. That is, no further refinement of the equations is necessary. The iterative solution process would involve shuffling the metric and inflaton perturbations back and forth between (4.4.25) and (4.4.1), alternately considering each as given in order to solve for the other.

---

[29] This agrees with [55, equation (5.6)]. It mostly agrees with the result on [60, page 179] except that Linde has used $\delta_k \approx -2\Phi_k$; when that approximation is used rather than the one used here in (4.4.21) then the results are identical.

[30] Compare to [60, equation (7.5.22)] but note how the errors Linde made in (1.7.17) etc. have compounded themselves.

[31] Compare this to the result in [60, equation (7.5.12)] and [69, equation (6.46)].

4.5 Comparison with Observation

4.5.1 Fluctuation Amplitude

In [55], Liddle and Lyth identify the quantity $\frac{4}{25}\mathcal{P}_{\mathcal{R}}(k)$ seen in (4.4.23) as

$$\frac{4}{25}\mathcal{P}_{\mathcal{R}}(k) = \delta_{\mathrm{H}}^2(k), \tag{4.5.1}$$

with $\delta_{\mathrm{H}}^2(k)$ being the mean square amplitude of density contrast fluctuations at horizon entry. This is a quantity that can be obtained from the COBE data set. The result of fitting the COBE data to a critical-density ($k = 0$) model universe with no gravitational waves and an assumed scale-invariant at horizon entry ($n_{\mathrm{s}\delta_{\mathrm{H}}} = 1$) spectrum gives[32]

$$\delta_{\mathrm{H}}(k) = 1.94 \times 10^{-5}. \tag{4.5.2}$$

Let us now compare our results to this observation. To do so, we'll start by writing $\delta_{\mathrm{H}}^2(k) = \frac{4}{25}\mathcal{P}_{\mathcal{R}}(k)$ explicitly in terms of $\phi|_{k\sim aH}$. Starting with

$$\delta_{\mathrm{H}}^2(k) = \frac{4}{25}\mathcal{P}_{\mathcal{R}}(k) = \frac{2\pi}{3}\left(\frac{16}{5}\right)^2 \left[\frac{V^3(\phi)}{V'^2(\phi)}\right]_{k\sim aH} \tag{4.5.3}$$

and assuming $V(\phi) = \alpha\phi^n$ we get

$$\delta_{\mathrm{H}}^2(k) = \frac{2\pi\alpha}{3}\left(\frac{16}{5n}\right)^2 \left[\phi^{n-2}\right]_{k\sim aH}. \tag{4.5.4}$$

Specifically, if $V(\phi) = \frac{1}{2}m^2\phi^2$ we have

$$\delta_{\mathrm{H}}^2(k) = \frac{\pi}{3}\left(\frac{8m}{5}\right)^2 \tag{4.5.5}$$

while if $V(\phi) = \frac{1}{4}\lambda\phi^4$ then

$$\delta_{\mathrm{H}}^2(k) = \frac{\pi\lambda}{6}\left(\frac{4}{5}\right)^2 \left[\phi^2\right]_{k\sim aH}. \tag{4.5.6}$$

Notice that an inflaton with $V(\phi) = \frac{1}{2}m^2\phi^2$ naturally gives not only a $n_{\mathrm{s}\delta_{\mathrm{H}}} =$ constant fluctuation spectrum but, specifically, a $n_{\mathrm{s}\delta_{\mathrm{H}}} = 1$ spectrum. The normalization of the spectrum in (4.5.2) determines the mass of the field which is, then,[33]

---

[32] See [55, Section 9.1.2].

$$m = 1.18 \times 10^{-5}. \tag{4.5.7}$$

The fact that the spectral index from a model like this is constant allows us to justify extending the fluctuation spectrum to scales well beyond those that can be accessed today through cosmological observations.

The $\lambda\phi^4$ model does not give a $n_{s\delta_H} = 1$ spectrum nor even a $n_{s\delta_H} = \text{constant}$ spectrum. To compare it with observation we need to find a way of specifying the value of the inflaton field, ϕ, when any mode, k, had a wavelength comparable in size to the de Sitter horizon. Let's obtain this information by asking the question in reverse: what is the current size of a mode that was the size of the horizon when the inflaton field had the value ϕ? From (2.3.19) it is seen that a mode will increase its wavelength by a factor of $\exp\left(\frac{4\pi}{n}\phi^2\right)$ from the time the inflaton field has a value of ϕ to the end of inflation. At the end of inflation, the universe reheats to a temperature T_{R} and subsequently cools to the temperature of the microwave background, T_γ, during which time it expands by another factor of T_{R}/T_γ [60, page 180]. Given all of this, the present wavelength of a mode whose wavelength was $H^{-1}(\phi)$ at a time when the inflaton field had a value of ϕ is

$$l(\phi) \sim H^{-1}(\phi)\frac{T_{\mathrm{R}}}{T_\gamma}\exp\left(\frac{4\pi}{n}\phi^2\right). \tag{4.5.8}$$

If we use (2.3.20) as the lower bound for ϕ during the inflation epoch, then for both $n = 2$ and $n = 4$ models we can say that inflation ends when $\phi \sim \frac{1}{3}$.[34] This sets the quantity of energy available for reheating. Prior to reheating the energy is in the inflaton field for which the energy goes as T^4 so $T_{\mathrm{R}} \sim \left[V\left(\frac{1}{3}\right)\right]^{1/4}$.[35] We'll take the final temperature to which the universe cools to be the current temperature of the CMBR, namely $T_\gamma \approx 3\ \mathrm{K} = 2 \times 10^{-32}$. Finally, using (2.3.13) for $H(\phi)$ and converting the result to centimetres gives

$$l(\phi) = \frac{1}{\sqrt{\frac{8\pi}{3}V(\phi)}}\frac{V^{\frac{1}{4}}\left(\frac{1}{3}\right)}{2 \times 10^{-32}}\exp\left(\frac{4\pi}{n}\phi^2\right)\left(1.6161 \times 10^{-33}\ \mathrm{cm}/l_{\mathrm{Pl}}\right). \tag{4.5.9}$$

For the $\frac{1}{4}\lambda\phi^4$ $(n = 4)$ model, this becomes

$$l(\phi) = \left(1.3 \times 10^{-2}\right)\lambda^{-\frac{1}{4}}\phi^{-2}\exp\left(\pi\phi^2\right) \tag{4.5.10}$$

or

$$\ln l(\phi) = -4.3 - \frac{1}{4}\ln\lambda - 2\ln\phi + \pi\phi^2. \tag{4.5.11}$$

---

[33] This is approximately an order of magnitude larger than the result Linde quotes in [60, equation (7.5.43)].

[34] This agrees with [60, pages 46 and 180].

[35] See [4, chapter 17, specifically 17.10] for details on the scaling of energy with temperature for this type of field.

The elimination of ϕ from (4.5.6) and (4.5.11) leads to a transcendental system of equations for λ as a function of $\delta_{\rm H}$ and $\ln l$. An approximate solution for λ can be had in the following way. We are going to compare our spectrum to structures on galactic ($l \sim 10^{23}$ cm) to horizon ($l \sim 10^{28}$ cm) distance scales — corresponding to $\ln l \in (53, 64)$. In this range of values, let us assume that the leading contribution to $\ln l$ comes from the $\pi\phi^2$ term and thus say that[36]

$$\ln l(\phi) \approx \pi\phi^2. \tag{4.5.12}$$

Using this as the expression for ϕ in (4.5.6) gives

$$\delta_{\rm H}^2 \approx \frac{\lambda}{6}\left(\frac{4}{5}\right)^2 \ln l \tag{4.5.13}$$

where l is in centimetres.[37] Note that for $53 < \ln l < 64$ (galactic to horizon scales), the predicted spectrum is flat to within 5% which is less than the uncertainty in the normalization. This approximate relationship between λ and $\delta_{\rm H}$ would lead to a value of $\lambda \approx 9 \times 10^{-10}$. The numerical solution of the transcendental equation gives a result of[38]

$$\lambda \approx 6 \times 10^{-11}. \tag{4.5.14}$$

The computations above set the magnitude of the inflaton field's parameters. By showing how the spectra are obtained from first principles, however, they also justify the extension of the observed perturbation spectra to scales larger than those that can be accessed through present day observations. This is very important for justifying the ability to make statements about the statistics of the collapse times of regions in the universe much larger than those that we can see.

---

[36] Compare to [60, equation 7.5.26].

[37] Compare to [60, equation7.5.29].

[38] This is between two and three orders of magnitude larger than the result quoted by Linde in [60, equation (7.5.37)].

Chapter 5

Collapse

Apart from being weak perturbations to the underlying FRW cosmos, the cosmological perturbations introduced in Chapters 3 and 4 have no restrictions on their structure. Analyzing the evolution of such arbitrary perturbations out of their linear phase to the point of their potential gravitational collapse is a very difficult problem. In order to analyze their behaviour further without yet resorting to numerical techniques, some further approximations will be made. Foremost among these will be the assumption that the perturbations are spherically symmetric, forming an onion-like structure around a preferred origin. It will be left until Section 5.2.1 to specify how this approximation will be made. For the time being, we will take it for granted that any perturbation variable, X, (for example the metric perturbation, Φ) is spherically symmetric, that is

$$X(t, r, \theta, \phi) = X(t, r).$$

With this restriction on the form of the perturbations, several approximation techniques for studying their long-term behaviour become available to us.

5.1 Spherically-Symmetric Model

5.1.1 Overview

In order to properly study the collapse time for a cosmological perturbation, we need a technique for evolving it out of the linear regime. We will accomplish this by introducing the Tolman-Lemaître description of cosmic structure. The Tolman-Lemaître metric tensor generalizes the RW tensor by allowing for a radially-dependant scale factor, $a(t, r)$, and curvature parameter, $k(r)$. This tensor is an exact solution to Einstein's equation and the time evolution of the scale factor can be obtained analytically when the source matter is assumed to be co-moving dust. The price to be paid for being able to obtain an exact solution is that our perturbations must be exactly spherically symmetric:

- the density field must pick out a preferred spatial origin about which it has no angular dependence,

- the cosmological fluid's velocity field must be perturbed only radially with respect to this origin, and
- the collapse time for the cosmological fluctuations must be a monotonically increasing function of radius.

The first condition is somewhat restrictive. The cosmological perturbation field has thusfar been described as a random field. While the statistics of the random field have been taken to be homogeneous and isotropic, any single realization of the perturbation field is an *inhomogeneous*, *anisotropic*, distribution of matter. It is extremely unlikely to find the entire field in a state of spherical symmetry. It is known, however, that the high peaks of a Gaussian random field are approximately spherically symmetric. This is shown in [7, Section VII] where the authors find that both the asphericity and the dispersion of peaks decreases in proportion to the inverse of the amplitude of the peaks — the vicinity of a larger peak is more spherically symmetric on average and more certain to be spherically symmetric. The conclusion is that the analysis to follow in this document can only be expected to be valid when describing the behaviour of the material in the vicinity of a high peak in the cosmological density field.

As for the second condition, if the density field is to be isotropic and the material is assumed to start out with no velocity perturbation at all, then velocity perturbations can only develop in the radial direction anyway. This second condition is equivalent to the statement that the perturbations have zero angular momentum.

The final condition is a consequence of the co-ordinate system being co-moving: we cannot allow matter from a larger radius to fall in on top of the matter at a smaller radius or the co-ordinate system will become singular. This condition will be not be addressed specifically in this document. For example, one must expect the imposition of this condition to influence the fluctuation spectrum, but here we will not make any modifications to our spectrum at all.

A few words should be said at this point about the applicability of this model to the observable universe. There are two issues to be addressed: (i) the assumption of the absence of pressure and (ii) the assumption of the absence of rotation. With regard to the former, it is known that even when the cosmological fluid is not pressureless, gravity more than pressure dominates the dynamics of structures larger than a particular length. This length is called the Jeans length and is given by [72, equation (4.8)]

$$L_{\mathrm{J}} = w\sqrt{\frac{\pi}{\rho}}$$

where $w = \mathrm{d}p/\,\mathrm{d}\rho$ describes the equation of state. From (2.1.14), in a flat universe the Jean's length is

$$L_{\mathrm{J}} = w\pi\sqrt{\frac{8}{3}}H^{-1}.$$

At the time of the decoupling of radiation from matter, the cosmological fluid was "radiation-like" so the equation of state was approximately $w = \frac{1}{3}$ and

$$L_{\mathrm{J\,dec}} = \frac{\pi}{3}\sqrt{\frac{8}{3}}H_{\mathrm{dec}}^{-1}. \tag{5.1.1}$$

In a flat, dust-filled universe, (2.1.13) tells us that $H \propto t^{-1}$, so

$$H_{\mathrm{dec}} = \frac{t_0}{t_{\mathrm{dec}}}H_0,$$

and from Table B.5, $t_0 = 13.7\pm0.2$ Ga, $t_{\mathrm{dec}} = 379^{+8}_{-7}$ ka, and $H_0 = 1.21\times10^{-61}\ {t_{\mathrm{Pl}}}^{-1}\pm3\%$ so

$$H_{\mathrm{dec}} = 4.37\times10^{-57}\ {t_{\mathrm{Pl}}}^{-1}\pm6\%, \tag{5.1.2}$$

and

$$L_{\mathrm{J\,dec}} \approx 3.9\times10^{56}\ l_{\mathrm{Pl}}. \tag{5.1.3}$$

(5.1.1) tells us that at the time of the decoupling of radiation from matter in a critical-density universe, a structure one Jean's-length across had a radius of

$$raH_{\mathrm{dec}} = \frac{1}{2}L_{\mathrm{J}}H_{\mathrm{dec}} = \frac{\pi}{6}\sqrt{\frac{8}{3}} \approx 0.8550, \tag{5.1.4}$$

so the collapse model is expected to be inapplicable to real structures in the universe below this radius.

If all of the material within a given radius collapses, then the Jean's length, which is inversely proportional to the energy density, is roughly proportional to the $\frac{3}{2}$ power of the material's radius — meaning that if gravitational forces dominate the dynamics of the material then they will always do so throughout its collapse and pressure can always be ignored. In general, however, one must expect the material's equation of state to change as it collapses which can lead to deviations from pure free-fall. For example, the material may condense into a swarm of point particles (e.g., stars) which "virialize" into orbits about the centre of mass. In this way, the ball of material can be stabilized against collapse by the angular momentum of its constituents, even if on the whole the material has no net angular momentum (half the particles are orbiting in one direction, and the other half in the other). In the case of this particular process it can be shown, and in general it is often assumed, that the halting of the collapse by "microphysics" happens very late in a collapsing ball of material's history.[1] Below, we will obtain the "collapse time" for a spherical ball of material. This is the time at which the material has shrunk to zero radius under gravitational free-fall. Although it is reasonable to expect some process to halt this collapse, and delay the formation of a black hole, such a process will not take over until very near the time of free-fall collapse. In this case, one may interpret the "collapse time" determined below to be the time at which microphysical processes, if present, dominate to produce some sort of compact system since the two times are very nearly one and the same. If the material ever shrinks inside its Schwarzschild radius, however, then gravitational collapse occurs regardless of the existence or otherwise of any microphysical process.

---

[1] See, for example [72, Section 8.2].

The second question to be addressed is the validity of the assumption of the lack of rotation in the collapsing material. Looking around in the universe, we see a wide variety of structures that have condensed out of the background expansion of the universe and nearly *all* of them are prevented from free-fall collapse into black holes through the support provided by angular momentum: for example, galaxies, binary neutron star systems, black hole accretion discs, and so on. Clearly rotation plays a significant role in determining the life-span of such systems. As mentioned in the previous paragraph, one can imagine angular momentum supporting a system against gravitational collapse even if the system as a whole has no net angular momentum. This is, in fact, the case for many of the larger of the examples listed above. By balancing gravitational and centrifugal forces, one can assign to a system the angular velocity the material within it would require in order for that material to be supported against gravitational collapse. The ratio of the actual observed net angular velocity to this gives a measure of the amount of rotational support available in a system. For elliptical galaxies, this is only ~ 0.05 while it can get as high as ~ 0.5 for spiral galaxies [72, Section 1.4]. For smaller systems, e.g. a binary neutron star system, this quantity is equal to 1 so one sees a trend of decreasing available rotational support with increasing size. In fact, it is generally assumed that at very large scales, there is no net angular momentum at all: the net rotation today observed in galaxies is taken to have originated from tidal torques among irregularly-shaped proto-galactic neighbours.[2] Additionally, during the period immediately following the decoupling of radiation from matter, the photon component of the cosmological fluid was still quite intense and it has been found that inverse Compton scattering of the photons off any circulating baryonic matter would have been an efficient mechanism for the dissipation of baryonic angular momentum [61, 91]. The assumption made here, that very large structures (e.g., larger than the Hubble radius at the time of matter-radiation decoupling) are devoid of any significant rotation, is consistent with current cosmological models. Any rotation that might be present, like the microphysical processes discussed in the previous paragraph, will be small and will not begin to significantly affect the collapse dynamics until very close to the end of the collapse process. In this case, once again the "collapse time" to be found below can be associated with the formation of, perhaps, a compact rotating system rather than a black hole.

5.1.2 Tolman-Lemaître Geometry

The standard Robertson-Walker geometry can be generalized to admit a scale factor that is a function of the radial co-ordinate as well as the time co-ordinate. Spherical symmetry in the spatial slices is maintained in this geometry but arbitrary radial inhomogeneities can be studied. The application of this geometry to cosmology was studied by Lemaître in [53] and by Tolman in [89]. A different approach to using this description of geometry to study the evolution of the cosmic density field can be found in [1]. The treatments that will be made use of here can be found in [74, Section 87] and [75, chapter 11]. If we consider the line element [74, equation (87.1)]

---

[2] See, for example [74, Section 23].

$$\mathrm{d}s^2 = -\,\mathrm{d}t^2 + \mathrm{e}^{2\alpha}\,\mathrm{d}r^2 + \mathrm{e}^{2\beta}\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right) \tag{5.1.5}$$

where $\alpha = \alpha(t,r)$ and $\beta = \beta(t,r)$, then Einstein's field equation results in four coupled differential equations for α and β in terms of T_{tt}, T_{rr}, $T_{\theta\theta}$ and T_{rt}. These can be found in [74, equation (87.2)] and are

$$8\pi T^0{}_0 = \dot\beta^2 + 2\dot\alpha\dot\beta + \mathrm{e}^{-2\beta} - \mathrm{e}^{-2\alpha}\left(2\beta'' + 3\beta'^2 - 2\alpha'\beta'\right) \tag{5.1.6a}$$

$$8\pi T^1{}_1 = 2\ddot\beta + 3\dot\beta^2 + \mathrm{e}^{-2\beta} - \beta'^2\mathrm{e}^{-2\alpha} \tag{5.1.6b}$$

$$8\pi T^2{}_2 = \ddot\alpha + \dot\alpha^2 + \ddot\beta + \dot\beta^2 + \dot\alpha\dot\beta - \mathrm{e}^{-2\alpha}\left(\beta'' + \beta'^2 - \alpha'\beta'\right) \tag{5.1.6c}$$

$$8\pi T^1{}_0 = 2\dot\beta' + 2\dot\beta\beta' - 2\dot\alpha\beta' \tag{5.1.6d}$$

where a dot represents differentiation with respect to t and a prime represents differentiation with respect to r.[3] The metric tensor described by the line element in (5.1.5) is the most general time-orthogonal, spherically-symmetric, metric tensor [74]. Time-orthogonal co-ordinates can always be found for a space-time geometry, although a single set of such co-ordinates may not be able to cover all space-time.

Imposing the requirement that the source matter be co-moving with the co-ordinates and be pressureless makes all components of T equal to 0 except T_{tt} which simplifies the dynamics considerably. In particular, it turns (5.1.6d) into a simple relationship between the time dependence of α and the time dependence of β

$$\frac{\dot\beta'}{\beta'} = \frac{\partial}{\partial t}\log\beta' = \dot\alpha - \dot\beta$$

which is easily integrated to give a relationship between β and α, namely[4]

$$\begin{aligned}\log\beta' &= \alpha - \beta + \log g(r)\\ \mathrm{e}^{\beta}\beta' &= \mathrm{e}^{\alpha}g(r)\\ \frac{\partial}{\partial r}\mathrm{e}^{\beta} &= \mathrm{e}^{\alpha}g(r).\end{aligned} \tag{5.1.7}$$

Writing e^{β} as[5]

$$\mathrm{e}^{\beta} = r a_{\mathrm{T}}(t,r) \tag{5.1.8}$$

and substituting into (5.1.7) gives

$$\left[r a_{\mathrm{T}}(t,r)\right]' = \mathrm{e}^{\alpha}g(r)$$

---

[3] Note that while Peebles is using a different sign convention, the differences do not affect the form of these equations.

[4] See [74, equation (87.4)] for this result.

[5] Compare to [74, equation (87.5)].

$$\mathrm{e}^{\alpha} = \frac{1}{g(r)}\,[a_{\mathrm{T}}(t,r)r]'$$

$$\mathrm{e}^{\alpha} = \frac{[a_{\mathrm{T}}(t,r)r]'}{\sqrt{1-k_{\mathrm{T}}(r)r^2}} \tag{5.1.9}$$

where $g(r)$ has been written as

$$g(r) = \sqrt{1-k_{\mathrm{T}}(r)r^2}. \tag{5.1.10}$$

Using (5.1.8) and (5.1.9) for e^{α} and e^{β} puts the line element in the form

$$\mathrm{d}s^2 = -\,\mathrm{d}t^2 + \frac{[a_{\mathrm{T}}r]'^2}{1-k_{\mathrm{T}}r^2}\,\mathrm{d}r^2 + {a_{\mathrm{T}}}^2r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right) \tag{5.1.11}$$

where $k_{\mathrm{T}} = k_{\mathrm{T}}(r)$ and $a_{\mathrm{T}} = a_{\mathrm{T}}(t,r)$. In the case that k_{T} and a_{T} are independent of r this reduces to (2.1.3), the Robertson-Walker line element. This is a nice property since it allows us to easily see how this geometry generalizes the pure isotropic homogeneous cosmological model. Because of the (intentional) similarities between the two line elements, here and in what follows, the notation X_{T} will be used to distinguish a quantity defined in the Tolman-Lemaître space-time from the corresponding quantity, X, in the Robertson-Walker space-time. For example, a is the Robertson-Walker scale factor while a_{T} is the Tolman-Lemaître scale factor.

The metric was put into the form given in (5.1.11) through the application of the constraint (5.1.6d) which means that in this form the metric automatically satisfies all off-diagonal parts of Einstein's equation. We have yet to impose (5.1.6a), (5.1.6b) or (5.1.6c). To determine the constraints these place on the solution, let us denote the radius of the 2-sphere part of the metric as R_{T}, i.e.

$$R_{\mathrm{T}}(t,r) = \mathrm{e}^{\beta} = a_{\mathrm{T}}(t,r)r. \tag{5.1.12}$$

From this,

$$\mathrm{e}^{\beta}\dot{\beta} = \dot{R}_{\mathrm{T}}$$

so

$$\dot{\beta} = \frac{\dot{R}_{\mathrm{T}}}{R_{\mathrm{T}}}, \qquad \ddot{\beta} = \frac{\ddot{R}_{\mathrm{T}}}{R_{\mathrm{T}}} - \dot{\beta}^2. \tag{5.1.13}$$

Similarly,

$$\beta' = \frac{{R_{\mathrm{T}}}'}{R_{\mathrm{T}}}, \qquad \beta'' = \frac{{R_{\mathrm{T}}}''}{R_{\mathrm{T}}} - \beta'^2. \tag{5.1.14}$$

Rewriting (5.1.9) as

$$\mathrm{e}^{\alpha} = \frac{{R_{\mathrm{T}}}'}{\sqrt{1-k_{\mathrm{T}}r^2}},$$

we get

$$e^{\alpha}\dot{\alpha} = \frac{\dot{R}_{\mathrm{T}}'}{\sqrt{1 - k_{\mathrm{T}}r^2}}$$

so

$$\dot{\alpha} = \frac{\dot{R}_{\mathrm{T}}'}{R_{\mathrm{T}}'}, \qquad \ddot{\alpha} = \frac{\ddot{R}_{\mathrm{T}}'}{R_{\mathrm{T}}'} - \dot{\alpha}^2. \tag{5.1.15}$$

Finally,

$$\begin{aligned} e^{\alpha}\alpha' &= \frac{R_{\mathrm{T}}''}{\sqrt{1 - k_{\mathrm{T}}r^2}} + \frac{1}{2}\frac{R_{\mathrm{T}}'\left(k_{\mathrm{T}}r^2\right)'}{\sqrt{1 - k_{\mathrm{T}}r^2}^3} \\ \alpha' &= \frac{R_{\mathrm{T}}''}{R_{\mathrm{T}}'} + \frac{1}{2}\frac{\left(k_{\mathrm{T}}r^2\right)'}{1 - k_{\mathrm{T}}r^2}. \end{aligned} \tag{5.1.16}$$

Using (5.1.13)–(5.1.16) in (5.1.6b) gives

$$\begin{aligned} 0 &= 2\frac{\ddot{R}_{\mathrm{T}}}{R_{\mathrm{T}}} + \left(\frac{\dot{R}_{\mathrm{T}}}{R_{\mathrm{T}}}\right)^2 + \frac{1}{R_{\mathrm{T}}^2} - \left(\frac{R_{\mathrm{T}}'}{R_{\mathrm{T}}}\right)^2 \frac{1 - k_{\mathrm{T}}r^2}{R_{\mathrm{T}}'^2} \\ &= 2R_{\mathrm{T}}\ddot{R}_{\mathrm{T}} + \dot{R}_{\mathrm{T}}^2 + k_{\mathrm{T}}r^2. \end{aligned} \tag{5.1.17}$$

Using (5.1.13)–(5.1.16) in (5.1.6c) gives

$$\begin{aligned} 0 &= \frac{\ddot{R}_{\mathrm{T}}'}{R_{\mathrm{T}}'} + \frac{\ddot{R}_{\mathrm{T}}}{R_{\mathrm{T}}} + \frac{\dot{R}_{\mathrm{T}}'}{R_{\mathrm{T}}'}\frac{\dot{R}_{\mathrm{T}}}{R_{\mathrm{T}}} - \frac{1 - k_{\mathrm{T}}r^2}{R_{\mathrm{T}}'^2}\left[\frac{R_{\mathrm{T}}''}{R_{\mathrm{T}}} - \left(\frac{R_{\mathrm{T}}''}{R_{\mathrm{T}}'} + \frac{1}{2}\frac{\left(k_{\mathrm{T}}r^2\right)'}{1 - k_{\mathrm{T}}r^2}\right)\frac{R_{\mathrm{T}}'}{R_{\mathrm{T}}}\right] \\ &= 2\ddot{R}_{\mathrm{T}}'R_{\mathrm{T}} + 2\ddot{R}_{\mathrm{T}}R_{\mathrm{T}}' + \dot{R}_{\mathrm{T}}'\dot{R}_{\mathrm{T}} + \left(k_{\mathrm{T}}r^2\right)' \end{aligned} \tag{5.1.18}$$

$$0 = \left(2R_{\mathrm{T}}\ddot{R}_{\mathrm{T}} + \dot{R}_{\mathrm{T}}^2 + k_{\mathrm{T}}r^2\right)' \tag{5.1.19}$$

which is simply (5.1.17) differentiated with respect to r and so one of (5.1.6b) and (5.1.6c) is redundant.

We will proceed by determining the behaviour of a shell at radius r. Substituting (5.1.8) and (5.1.9) into (5.1.6a) and (5.1.6b) gives[6]

$$\frac{8\pi}{3}\rho\frac{\partial}{\partial r}\left(a_{\mathrm{T}}r\right)^3 = \frac{\partial}{\partial r}\left(\dot{a}_{\mathrm{T}}^2 a_{\mathrm{T}}r^3 + k_{\mathrm{T}}a_{\mathrm{T}}r^3\right) \tag{5.1.20a}$$

$$2\frac{\ddot{a}_{\mathrm{T}}}{a_{\mathrm{T}}} + \left(\frac{\dot{a}_{\mathrm{T}}}{a_{\mathrm{T}}}\right)^2 + \frac{k_{\mathrm{T}}}{a_{\mathrm{T}}^2} = 0. \tag{5.1.20b}$$

---

[6] An equivalent result can be found in [74, (87.7)].

Multiplying (5.1.20b) by ${a_{\mathrm{T}}}^2\dot{a}_{\mathrm{T}}$ puts it in the form

$$\frac{\partial}{\partial t_{\mathrm{T}}}\left(\dot{a}_{\mathrm{T}}^2 a_{\mathrm{T}} + k_{\mathrm{T}} a_{\mathrm{T}}\right) = 0$$

which, integrated, is

$$\dot{a}_{\mathrm{T}}^2 a_{\mathrm{T}} + k_{\mathrm{T}} a_{\mathrm{T}} = F(r). \tag{5.1.21}$$

We need to solve (5.1.21) separately for radii where $k_{\mathrm{T}}(r) < 0$, $k_{\mathrm{T}}(r) = 0$ and $k_{\mathrm{T}}(r) > 0$.

For radii with $k_{\mathrm{T}}(r) < 0$, (5.1.21) can be solved parametrically. Expressing a_{T} and t_{T} in terms of a third quantity, η_{T}, with $\frac{\mathrm{d}a_{\mathrm{T}}}{\mathrm{d}t_{\mathrm{T}}} = \frac{\mathrm{d}a_{\mathrm{T}}}{\mathrm{d}\eta_{\mathrm{T}}}\frac{\mathrm{d}\eta_{\mathrm{T}}}{\mathrm{d}t_{\mathrm{T}}}$ we get

$$\left(\frac{\mathrm{d}a_{\mathrm{T}}}{\mathrm{d}\eta_{\mathrm{T}}}\right)^2 a_{\mathrm{T}} + \left(\frac{\mathrm{d}t_{\mathrm{T}}}{\mathrm{d}\eta_{\mathrm{T}}}\right)^2 a_{\mathrm{T}} k_{\mathrm{T}} = \left(\frac{\mathrm{d}t_{\mathrm{T}}}{\mathrm{d}\eta_{\mathrm{T}}}\right)^2 F$$

and assuming separable solutions with

$$a_{\mathrm{T}}(r, \eta_{\mathrm{T}}) = \frac{F(r)}{k_{\mathrm{T}}(r)} A(\eta_{\mathrm{T}})$$

$$t_{\mathrm{T}}(\eta_{\mathrm{T}}) = \frac{F(r)}{\left|k_{\mathrm{T}}\right|^{3/2}(r)} T(\eta_{\mathrm{T}})$$

gives

$$\left(\frac{\mathrm{d}A}{\mathrm{d}\eta_{\mathrm{T}}}\right)^2 A - \left(\frac{\mathrm{d}T}{\mathrm{d}\eta_{\mathrm{T}}}\right)^2 A = \left(\frac{\mathrm{d}T}{\mathrm{d}\eta_{\mathrm{T}}}\right)^2.$$

It is easily checked that the solutions for A and T are

$$A(\eta_{\mathrm{T}}) = \frac{1}{2}\left(\cosh\eta_{\mathrm{T}} - 1\right)$$

$$T(\eta_{\mathrm{T}}) = \frac{1}{2}\left(\sinh\eta_{\mathrm{T}} - \eta_{\mathrm{T}}\right) + T_0$$

so, finally, the parametric expressions for a_{T} and t_{T} are

$$a_{\mathrm{T}}(r, \eta_{\mathrm{T}}) = \frac{F(r)}{2k_{\mathrm{T}}(r)}\left(\cosh\eta_{\mathrm{T}} - 1\right) \tag{5.1.22a}$$

$$t_{\mathrm{T}}(r, \eta_{\mathrm{T}}) = \frac{F(r)}{2\left|k_{\mathrm{T}}\right|^{3/2}(r)}\left(\sinh\eta_{\mathrm{T}} - \eta_{\mathrm{T}}\right) + t_{\mathrm{T}0} \tag{5.1.22b}$$

with $\eta_{\mathrm{T}} \in [0, \infty)$. This indicates that shells at radii for which $k_{\mathrm{T}}(r) < 0$ do not experience collapse.

For radii with $k_{\mathrm{T}}(r) = 0$, (5.1.21) can be integrated to give

$$a_{\mathrm{T}}(r, t_{\mathrm{T}}) = \left[\frac{9}{4}F(r)\right]^{\frac{1}{3}}\left[\pm t_{\mathrm{T}} - t_{\mathrm{T}0}(r)\right]^{\frac{2}{3}}. \tag{5.1.23}$$

Algebraically, then, these radii are either expanding monotonically ($+t_{\mathrm{T}}$ case) or collapsing monotonically ($-t_{\mathrm{T}}$ case). Although radii for which $k_{\mathrm{T}} = 0$ may be collapsing, the case of $k_{\mathrm{T}} = 0$ represents a measure 0 fraction of the parameter space. To simplify the following analysis, these radii will be discarded from the set of "collapsing radii."

For radii with $k_{\mathrm{T}}(r) > 0$, (5.1.21) can be solved parametrically. Again, expressing a_{T} and t_{T} in terms of a third quantity, η_{T}, with $\frac{\mathrm{d}a_{\mathrm{T}}}{\mathrm{d}t_{\mathrm{T}}} = \frac{\mathrm{d}a_{\mathrm{T}}}{\mathrm{d}\eta_{\mathrm{T}}}\frac{\mathrm{d}\eta_{\mathrm{T}}}{\mathrm{d}t_{\mathrm{T}}}$ we get

$$\left(\frac{\mathrm{d}a_{\mathrm{T}}}{\mathrm{d}\eta_{\mathrm{T}}}\right)^2 a_{\mathrm{T}} + \left(\frac{\mathrm{d}t_{\mathrm{T}}}{\mathrm{d}\eta_{\mathrm{T}}}\right)^2 a_{\mathrm{T}}k_{\mathrm{T}} = \left(\frac{\mathrm{d}t_{\mathrm{T}}}{\mathrm{d}\eta_{\mathrm{T}}}\right)^2 F$$

and, again, assuming separable solutions with

$$a_{\mathrm{T}}(r,\eta_{\mathrm{T}}) = \frac{F(r)}{k_{\mathrm{T}}(r)}A(\eta_{\mathrm{T}})$$
$$t_{\mathrm{T}}(\eta_{\mathrm{T}}) = \frac{F(r)}{|k_{\mathrm{T}}|^{3/2}(r)}T(\eta_{\mathrm{T}})$$

gives[7]

$$\left(\frac{\mathrm{d}A}{\mathrm{d}\eta_{\mathrm{T}}}\right)^2 A + \left(\frac{\mathrm{d}T}{\mathrm{d}\eta_{\mathrm{T}}}\right)^2 A = \left(\frac{\mathrm{d}T}{\mathrm{d}\eta_{\mathrm{T}}}\right)^2.$$

It is easily checked that the solutions for A and T are

$$A(\eta_{\mathrm{T}}) = \frac{1}{2}\left(1 - \cos\eta_{\mathrm{T}}\right)$$
$$T(\eta_{\mathrm{T}}) = \frac{1}{2}\left(\eta_{\mathrm{T}} - \sin\eta_{\mathrm{T}}\right) + T_0$$

so, finally, the parametric expressions for a_{T} and t_{T} are

$$a_{\mathrm{T}}(r,\eta_{\mathrm{T}}) = \frac{F(r)}{2k_{\mathrm{T}}(r)}\left(1 - \cos\eta_{\mathrm{T}}\right) \tag{5.1.24a}$$
$$t_{\mathrm{T}}(r,\eta_{\mathrm{T}}) = \frac{F(r)}{2\,|k_{\mathrm{T}}|^{3/2}(r)}\left(\eta_{\mathrm{T}} - \sin\eta_{\mathrm{T}}\right) + t_{\mathrm{T}0} \tag{5.1.24b}$$

with $\eta_{\mathrm{T}} \in [0, 2\pi]$. This indicates that shells at radii for which $k_{\mathrm{T}}(r) > 0$ do experience collapse. To determine the time required for the collapse we first need to find a way to determine the value of η_{T} that corresponds to each point in the space-time. Eliminating F/a_{T} from (5.1.21) and (5.1.24a), then solving for η_{T} gives

$$\eta_{\mathrm{T}}(r,t_{\mathrm{T}}) = 2\tan^{-1}\frac{\sqrt{k_{\mathrm{T}}(r)}}{\dot{a}_{\mathrm{T}}(r,t_{\mathrm{T}})}. \tag{5.1.25}$$

This expression gives the value of η_{T} at each point in the space-time if given the values of $k_{\mathrm{T}}(r)$ and $\dot{a}_{\mathrm{T}}(r,t_{\mathrm{T}})$ at that point. The collapsed singularity occurs at $\eta_{\mathrm{T}} = 2\pi$, so from

---

[7] Note the $+$ in this equation that was a $-$ for the $k_{\mathrm{T}} < 0$ case.

(5.1.24b) the proper time which elapses from any given (doomed) point to its eventual collapse is given by

$$\begin{aligned} t_{\text{collapse}}(r) &= \frac{F}{{k_{\mathrm{T}}}^{3/2}}\left[\pi - \tan^{-1}\frac{\sqrt{k_{\mathrm{T}}}}{\dot{a}_{\mathrm{T}}} + \frac{1}{2}\sin\tan^{-1}\frac{\sqrt{k_{\mathrm{T}}}}{\dot{a}_{\mathrm{T}}}\right] \\ &= \frac{\dot{a}_{\mathrm{T}}^2}{k_{\mathrm{T}}}\sqrt{\frac{{a_{\mathrm{T}}}^2}{k_{\mathrm{T}}}}\left[1 + \frac{k_{\mathrm{T}}}{\dot{a}_{\mathrm{T}}^2}\right]\left[\pi - \tan^{-1}\frac{\sqrt{k_{\mathrm{T}}}}{\dot{a}_{\mathrm{T}}} + \frac{1}{2}\frac{\sqrt{k_{\mathrm{T}}}}{\dot{a}_{\mathrm{T}}}\left(1 + \frac{k_{\mathrm{T}}}{\dot{a}_{\mathrm{T}}^2}\right)^{-\frac{1}{2}}\right] \end{aligned} \tag{5.1.26}$$

where (5.1.21) has been used as an expression for the integration constant, $F(r)$.

From (5.1.26) it is seen that the collapse time depends only on the quantities $\frac{k_{\mathrm{T}}}{{a_{\mathrm{T}}}^2}$ and $\frac{k_{\mathrm{T}}}{\dot{a}_{\mathrm{T}}^2}$. These are, essentially, the initial-value data for the collapse model. In Sections 5.1.3 and 5.1.4 below, we will obtain expressions for these quantities on a space-like hyper-surface from the parameters of the conformal Newtonian description of geometry on the same surface. Those results can be used directly in (5.1.26), however the final expression need only given to first order in the Newtonian perturbations, Φ. It will be found that k_{T} is functionally proportional to the Newtonian perturbations and with this foreknowledge of the results, it is possible to perform some further simplifications at this time. Since k_{T} is functionally proportional to Φ, the quantities in both sets of brackets in (5.1.26) need only be evaluated to zeroth order in k_{T}. This leaves us with simply

$$t_{\text{collapse}}(r) \approx \pi\frac{\dot{a}_{\mathrm{T}}^2}{k_{\mathrm{T}}}\sqrt{\frac{{a_{\mathrm{T}}}^2}{k_{\mathrm{T}}}}. \tag{5.1.27}$$

To use this result to determine the collapse time for some structure we now need to translate our perturbations into the Tolman-Lemaître description of the universe's geometry. Reviewing, we consider the conformal Newtonian description of the geometry of space-time to only be applicable at early times, when linear perturbation theory is accurate. At late times, when structures are evolving out of their linear regime, the fully non-linear Tolman-Lemaître description of geometry must be used. The result of the analysis of the evolution of structures in the framework of the conformal Newtonian model is a scalar perturbation field, Φ. We wish to construct appropriate initial conditions for the Tolman-Lemaître collapse model from knowledge of Φ. We shall do this (even if only conceptually) by choosing a space-like hyper-surface to act as a boundary between the two models: prior to the epoch of this hyper-surface, the geometry of space-time is described by the scalar field Φ; following the epoch of this hyper-surface, the geometry of space-time is to be described by the parameters of the Tolman-Lemaître model. We shall use two quite different techniques for extracting initial conditions for the Tolman-Lemaître model from the Newtonian model in order to check the consistency of the results.

It must be stressed that we will be obtaining *initial conditions* for the Tolman-Leaître model. The conclusion of the procedures by which this is done is an expression for the collapse time in (3.4.19) re-expressed in terms of the Newtonian parameters. One must not think this means that the collapse process is being or can be described by the

Newtonian model. The collapse is described by the Tolman-Lemaître model. Since the Newtonian line element is only an approximate solution of Einstein's equation while the Tolman-Lemaître line element is an exact one, it is not possible for the two descriptions of geometry to agree everywhere. Although, as will be shown below, they can be brought into agreement on any single space-like hyper-surface, it must be expected that as they are evolved away from the chosen boundary surface they will fall into disagreement, though the disagreement should be small during the epoch that is well approximated by the Newtonian description. Finally, although any single space-like hyper-surface can be chosen to act as the boundary between the two descriptions of geometry, it is important that the particular hyper-surface chosen occurs during the epoch that the Newtonian description is accurate.

5.1.3 Construction of $t_{\mathrm{collapse}}(r)$ From Invariant Scalar Expressions

The technique used to obtain $k_{\mathrm{T}}(r_{\mathrm{T}})$, $a_{\mathrm{T}}(r_{\mathrm{T}}, t_{\mathrm{T}})$, and $\dot{a}_{\mathrm{T}}(r_{\mathrm{T}}, t_{\mathrm{T}})$ in this section will be to discover a way of constructing these quantities from scalar fields computed from the Tolman-Lemaître metric tensor. Applying the same construction to the conformal Newtonian metric tensor will then give us the values of these quantities at any given event in the Newtonian co-ordinate system. As mentioned at the end of the previous section, the collapse time depends only on the quantities $\frac{k_{\mathrm{T}}}{a_{\mathrm{T}}{}^2}$ and $\frac{k_{\mathrm{T}}}{\dot{a}_{\mathrm{T}}^2}$. If we can construct invariant expressions for these from the Tolman-Lemaître metric tensor then we have solved the problem.

In both the Newtonian and Tolman line-elements, the geometry at each point is expressed in terms of a 2-sphere (the angular co-ordinates) crossed with a time-space 2-plane and in both cases we can readily identify the quantity corresponding to the radius of the 2-sphere. Let's call this radius R. The value of R at each space-time event is given, in the Tolman case, by (5.1.12). In the Newtonian case we can obtain it from the line element in (4.3.20) and it is

$$\begin{aligned} R^2(t,r) &= [1 - 2\Phi(t,r)]\, a^2(t) r^2 \\ R(t,r) &\approx [1 - \Phi(t,r)]\, a(t) r. \end{aligned} \tag{5.1.28}$$

Henceforth we will drop the distinction between the Tolman R_{T} and the Newtonian R by setting

$$R = R_{\mathrm{T}}. \tag{5.1.29}$$

One should remember that the functional form of the two radii differ with regard to their respective co-ordinate systems. This is not important, however; it is only important that we know how to determine R in one co-ordinate system given R in the other and this we now know how to do. Let us continue by examining the covariant derivatives of R with respect to the t-r 2-metric in the Tolman geometry to see if we can assemble the pieces needed to construct (5.1.27). The first scalar invariant we can compute from the first derivatives of R is $R^{;A}R_{;A}$ which, in the Tolman case, is

$$\begin{aligned} R^{;A}R_{;A} &= g^{AB}R_{,A}R_{,B} \\ &= -\dot{R}^2 + \frac{1-k_{\mathrm{T}}r^2}{R'^2}R'^2 \\ R^{;A}R_{;A} &= 1 - k_{\mathrm{T}}r^2 - \dot{R}^2. \end{aligned} \tag{5.1.30}$$

Now $R^{;A}R_{;A}$ is itself a scalar quantity and from it let's define a third scalar,

$$M = \frac{1}{2}R\left(1 - R^{;A}R_{;A}\right). \tag{5.1.31}$$

In the Tolman case, using (5.1.12) we find that

$$M = \frac{1}{2}R\left(k_{\mathrm{T}}r^2 + \dot{R}^2\right). \tag{5.1.32}$$

Before continuing, note that due to the constraint in (5.1.17),

$$\dot{M} = \frac{1}{2}\dot{R}\left(k_{\mathrm{T}}r^2 + 2\ddot{R}R + \dot{R}^2\right) = 0. \tag{5.1.33}$$

Applying (5.1.33) in the computation of invariants constructed from the derivatives of M we find that

$$\begin{aligned} M^{;A}M_{;A} &= g^{AB}M_{,A}M_{,B} \\ &= -\dot{M}^2 + \frac{1-k_{\mathrm{T}}r^2}{R'^2}M'^2 \\ M^{;A}M_{;A} &= \frac{1-k_{\mathrm{T}}r^2}{R'^2}M'^2 \end{aligned} \tag{5.1.34}$$

and

$$\begin{aligned} R^{;A}M_{;A} &= g^{AB}R_{,A}M_{,B} \\ &= -\dot{R}\dot{M} + \frac{1-k_{\mathrm{T}}r^2}{R'^2}R'M' \\ R^{;A}M_{;A} &= \frac{1-k_{\mathrm{T}}r^2}{R'}M'. \end{aligned} \tag{5.1.35}$$

Squaring (5.1.35) and dividing by (5.1.34) allows us to solve for $k_{\mathrm{T}}r^2$ and gives

$$k_{\mathrm{T}}r^2 = 1 - \frac{\left(R^{;A}M_{;A}\right)^2}{M^{;A}M_{;A}} \tag{5.1.36}$$

where the right-hand side is a fully invariant expression. Substituting (5.1.36) into (5.1.30) gives an invariant expression that is equivalent to $\dot{R}^2$ in the Tolman geometry, namely

$$\dot{R}^2 = \frac{\left(R^{;A}M_{;A}\right)^2}{M^{;A}M_{;A}} - R^{;A}R_{;A}. \tag{5.1.37}$$

Finally, alternately dividing the invariant expression for $k_{\mathrm{T}}r^2$ in (5.1.36) by R^2 in (5.1.12) and $\dot{R}^2$ in (5.1.37) gives us our desired results, namely

$$\frac{k_{\mathrm{T}}}{a^2} = \frac{k_{\mathrm{T}}r^2}{R^2} = \frac{1}{R^2}\left[1 - \frac{\left(R^{;A}M_{;A}\right)^2}{M^{;A}M_{;A}}\right], \tag{5.1.38}$$

and

$$\frac{k_{\mathrm{T}}}{\dot{a}^2} = \frac{k_{\mathrm{T}}r^2}{\dot{R}^2} = \left[1 - \frac{\left(R^{;A}M_{;A}\right)^2}{M^{;A}M_{;A}}\right]\left[\frac{\left(R^{;A}M_{;A}\right)^2}{M^{;A}M_{;A}} - R^{;A}R_{;A}\right]^{-1}. \tag{5.1.39}$$

Substituting (5.1.38) and (5.1.39) into the expression for the collapse time in (5.1.27) gives us the collapse time for a shell at radius r solely in terms of local invariant quantities computed from the line element,

$$\begin{aligned} t_{\mathrm{collapse}}(r) &\approx \pi \frac{\dot{a}_{\mathrm{T}}^2}{k_{\mathrm{T}}}\sqrt{\frac{a_{\mathrm{T}}{}^2}{k_{\mathrm{T}}}} \\ &= \pi R\left[\frac{\left(R^{;A}M_{;A}\right)^2}{M^{;A}M_{;A}} - R^{;A}R_{;A}\right]\left[1 - \frac{\left(R^{;A}M_{;A}\right)^2}{M^{;A}M_{;A}}\right]^{-\frac{3}{2}}. \end{aligned} \tag{5.1.40}$$

We now have all the pieces required to compute $t_{\mathrm{collapse}}(r)$ in the Newtonian description of space-time geometry. Proceeding from the Newtonian line element, (4.3.19), and keeping results accurate to only first order in the perturbations we begin by finding

$$\begin{aligned} R^{;A}R_{;A} &= g^{AB}R_{,A}R_{,B} \\ &= -\frac{1}{1+2\Phi}\dot{R}^2 + \frac{1}{(1+2\Phi)a^2}R'^2 \\ &\approx -\left[(1-\Phi)\,\dot{R}\right]^2 + \left[(1+\Phi)\,a^{-1}R'\right]^2 \qquad (5.1.41) \\ &= -\left\{[1-\Phi]\left[(1-\Phi)\dot{a}r - \dot{\Phi}ar\right]\right\}^2 + \left\{[1+\Phi]\,a^{-1}\left[(1-\Phi)a - \Phi' ar\right]\right\}^2 \\ R^{;A}R_{;A} &\approx -a^2r^2\frac{\dot{a}}{a}\left[(1-4\Phi)\,\frac{\dot{a}}{a} - 2\dot{\Phi}\right] + 1 - 2\Phi' r. \qquad (5.1.42) \end{aligned}$$

This allows the computation of M which, according to its definition in (5.1.31), is

$$M = \frac{1}{2}\,(1-\Phi)\,ar\left\{a^2r^2\frac{\dot{a}}{a}\left[(1-4\Phi)\,\frac{\dot{a}}{a} - 2\dot{\Phi}\right] + 2\Phi' r\right\}$$

$$\approx \frac{1}{2}a^3r^3\frac{\dot{a}}{a}\left[(1-5\Phi)\frac{\dot{a}}{a}-2\dot{\Phi}\right]+\Phi' ar^2. \tag{5.1.43}$$

At this point, it proves fruitful to investigate $\dot{M}$ which is

$$\begin{aligned}\dot{M} &= \frac{1}{2}\left(3a^2\dot{a}r^3\frac{\dot{a}}{a}+a^3r^3\frac{\ddot{a}a-\dot{a}^2}{a^2}\right)\left[(1-5\Phi)\frac{\dot{a}}{a}-2\dot{\Phi}\right]+\\ &\qquad \frac{1}{2}a^3r^3\frac{\dot{a}}{a}\left[(1-5\Phi)\frac{\ddot{a}a-\dot{a}^2}{a^2}-5\dot{\Phi}\frac{\dot{a}}{a}-2\ddot{\Phi}\right]+\Phi'\dot{a}r^2+\dot{\Phi}'ar^2\\ \dot{M} &= -a^3r^3\frac{\dot{a}}{a}\left(4\dot{\Phi}\frac{\dot{a}}{a}+\ddot{\Phi}\right)+ar^2\left(\Phi'\frac{\dot{a}}{a}+\dot{\Phi}'\right). \end{aligned} \tag{5.1.44}$$

where one of the equations of motion for the background $k = 0$ pressureless FRW universe, namely (2.1.6a), has been applied as a constraint. This result can be simplified further by applying the equations of motion for Φ during the dust-filled FRW phase of the universe. These are given in Section C.9.2. In particular, using (C.9.36c) gives us

$$\dot{M} = ar^2\left(\Phi'\frac{\dot{a}}{a}+\dot{\Phi}'\right), \tag{5.1.45}$$

Notice that in the limit where there are no perturbations, $\dot{M} = 0$ just as in the Tolman case. Also, since to zeroth order in the perturbations $\dot{M} = 0$, to first order in the perturbations $\dot{M}^2 = 0$. These properties of $\dot{M}$ allow us to perform some simplifications of the expression for the collapse time, (5.1.40), before proceeding. In particular, we find that

$$M^{;A}M_{;A} = -\frac{1}{1+2\Phi}\dot{M}^2+\frac{1}{(1-2\Phi)\,a^2}M'^2 \approx \left[(1+\Phi)\,a^{-1}M'\right]^2, \tag{5.1.46}$$

and

$$R^{;A}M_{;A} = -\frac{1}{1+2\Phi}\dot{R}\dot{M}+\frac{1}{(1-2\Phi)\,a^2}R'M' \approx \left[(1+\Phi)\,a^{-1}R'\right]\left[(1+\Phi)\,a^{-1}M'\right]-\dot{R}\dot{M}. \tag{5.1.47}$$

Combining (5.1.46) with (5.1.47), and also using $\dot{M}^2 \approx 0$, we find that

$$\begin{aligned}\frac{\left(R^{;A}M_{;A}\right)^2}{M^{;A}M_{;A}} &\approx \left[(1+\Phi)\,a^{-1}M'\right]^{-2}\left[(1+\Phi)\,a^{-1}R'\right]\left[(1+\Phi)\,a^{-1}M'\right]\times\\ &\qquad \left\{\left[(1+\Phi)\,a^{-1}R'\right]\left[(1+\Phi)\,a^{-1}M'\right]-2\dot{R}\dot{M}\right\}\\ &= \left[(1+\Phi)\,a^{-1}R'\right]^2-2\dot{R}\dot{M}\frac{R'}{M'}. \end{aligned} \tag{5.1.48}$$

Finally, using (5.1.48) along with the structural form of $R^{;A}R_{;A}$ found in (5.1.41), we find that

$$t_{\text{collapse}}(r) = \pi R\left[\frac{\left(R^{;A}M_{;A}\right)^2}{M^{;A}M_{;A}} - R^{;A}R_{;A}\right]\left[1-\frac{\left(R^{;A}M_{;A}\right)^2}{M^{;A}M_{;A}}\right]^{-\frac{3}{2}} \tag{5.1.49}$$

$$= \pi R\left\{\left[\left(1-\Phi\right)\dot{R}\right]^2 - 2\dot{R}\dot{M}\frac{R'}{M'}\right\}\left\{1-\left[\left(1+\Phi\right)a^{-1}R'\right]^2 + 2\dot{R}\dot{M}\frac{R'}{M'}\right\}^{-\frac{3}{2}}$$

$$t_{\text{collapse}}(r) \approx \pi a^2r^2\left\{ar\frac{\dot{a}}{a}\left[\left(1-5\Phi\right)\frac{\dot{a}}{a} - 2\dot{\Phi}\right] - 2\dot{a}\frac{\dot{M}}{M'}\right\}\left\{2\Phi' r + 2a\dot{a}r\frac{\dot{M}}{M'}\right\}^{-\frac{3}{2}}. \tag{5.1.50}$$

All that remains is to compute M'. In doing this, one should note that in fact we only need the quantity $\dot{M}/M'$ and since $\dot{M} = 0$ to zeroth order in the perturbations, we only need to compute M' to zeroth order. Linear and higher order terms in M' would result in quadratic and higher terms when M'^{-1} is multiplied by $\dot{M}$ and so are not needed. This line of approximation goes further, however. Since $\dot{M}/M' = 0$ to zeroth order in the perturbations, notice that the entire denominator in (5.1.50) is equal to 0 to zeroth order in the perturbations. This means that first order and higher terms can be discarded from the numerator so, in fact,

$$t_{\text{collapse}}(r) \approx \pi\left(ar\right)^3\left(\frac{\dot{a}}{a}\right)^2\left\{2\Phi' r + 2\left(ar\right)\dot{a}\frac{\dot{M}}{M'}\right\}^{-\frac{3}{2}}. \tag{5.1.51}$$

Proceeding with the computation of $\dot{M}/M'$ we get the following for M' to zeroth order,

$$M' = \frac{1}{2}3a^3r^2\frac{\dot{a}}{a}\left[\left(1-5\Phi\right)\frac{\dot{a}}{a} - 2\dot{\Phi}\right] + \frac{1}{2}a^3r^3\frac{\dot{a}}{a}\left[-5\Phi'\frac{\dot{a}}{a} - 2\dot{\Phi}'\right] + \Phi'' ar^2 + 2\Phi' ar$$

$$\approx \frac{3}{2}a^3r^2\left(\frac{\dot{a}}{a}\right)^2. \tag{5.1.52}$$

This could, of course, have been computed directly from the zeroth order term in M but this shows the full form of M' to linear order. Combining this result with the expression for $\dot{M}$ in (5.1.45) gives

$$\frac{\dot{M}}{M'} = \left[2ar^2\left(\Phi'\frac{\dot{a}}{a} + \dot{\Phi}'\right)\right]\left[3a^3r^2\left(\frac{\dot{a}}{a}\right)^2\right]^{-1} = \frac{2}{3a^2}\left(\frac{\dot{a}}{a}\right)^{-2}\left(\Phi'\frac{\dot{a}}{a} + \dot{\Phi}'\right). \tag{5.1.53}$$

Using this in (5.1.51) gives

$$t_{\text{collapse}}(r) \approx \pi\left(ra\right)^3\left(\frac{\dot{a}}{a}\right)^2\left\{2\Phi' r + 2\left(ra\right)\dot{a}\frac{\dot{M}}{M'}\right\}^{-\frac{3}{2}}$$

$$= \frac{\pi}{\sqrt{2r}^3} (ra)^3 \left(\frac{\dot{a}}{a}\right)^2 \left\{ \Phi' + \frac{2}{3} \left(\frac{\dot{a}}{a}\right)^{-1} \left(\Phi' \frac{\dot{a}}{a} + \dot{\Phi}' \right) \right\}^{-\frac{3}{2}}$$

$$t_{\text{collapse}}(r) = \frac{\pi}{\sqrt{2}^3} (raH)^{\frac{3}{2}} H^{-1} \left(\frac{\Phi'}{aH} - a\delta u_r \right)^{-\frac{3}{2}} . \tag{5.1.54}$$

In obtaining this result we have used (C.9.36b), one of the equations of motion for the metric perturbations when the cosmological fluid is dust. $a\delta u_r$ is the proper radial velocity of the fluid as observed in the cosmological rest frame.

Let us determine the condition that must be satisfied in order for a shell at the radius r to collapse. As discussed above in Section 5.1.2, in order for a shell to collapse we must have $k_{\text{T}} > 0$. From (5.1.36) we see that this is equivalent to requiring

$$1 - \frac{\left(R^{;A} M_{;A}\right)^2}{M^{;A} M_{;A}} > 0.$$

This is a convenient condition as the left-hand side is simply the expression that appears as the base in the numerator of $t_{\text{collapse}}(r)$ in (5.1.49). In other words, the condition that a shell actually collapse can also be stated as simply

$$t_{\text{collapse}}(r) \in \text{Reals}. \tag{5.1.55}$$

5.1.4 Construction of $t_{\text{collapse}}(r)$ By Geometry Matching

The second technique we will use to construct t_{collapse} is to explicitly stitch the Newtonian and Tolman geometries together along a space-like interface hyper-surface. In effect, the technique is to use the late-time spatial geometry arrived at using the Newtonian perturbation analysis as the initial conditions for the Tolman solution.

The issues involved in connecting two solutions of Einstein's equation together are covered in [37] and [65, Section 21.13]. Conceptually the issues are very similar to those of interface conditions in electro-dynamics: if there are no surface sources of curvature sitting on the boundary, then the geometry of space-time must be smooth across it. In other words, both the intrinsic and extrinsic geometries of the boundary hyper-surface in the first solution must be identical to those quantities for the boundary hyper-surface in the second solution. The mechanics of the procedure for enforcing these equalities is complicated by the fact that we are free to choose how it is we identify co-ordinates on one side of the boundary with co-ordinates on the other. This co-ordinate mapping must also be supplied as part of the matching conditions.

Let's begin by considering the continuity of the intrinsic geometry and the co-ordinate mapping. Our intention is for the geometry described by the Tolman-Lemaître metric tensor in (5.1.11) to be exactly the same as the geometry described by the conformal Newtonian metric tensor in (4.3.20); in particular on the boundary surface. In other words, we want them to be the same tensor. Being the same tensor, they are related by the standard co-ordinate transformation rule for second-rank covariant tensors, namely

$$g_{\mathrm{T}\,\mu_{\mathrm{T}}\nu_{\mathrm{T}}} = \frac{\mathrm{d}x^{\alpha}}{\mathrm{d}x_{\mathrm{T}}{}^{\mu}} \frac{\mathrm{d}x^{\beta}}{\mathrm{d}x_{\mathrm{T}}{}^{\nu}} g_{\alpha\beta}$$

where a subscript T indicates the Tolman-Lemaître version of the particular quantity (metric tensor, co-ordinate, etc.). The Newtonian metric tensor is an accurate description of geometry to only first order in the metric perturbations so this co-ordinate transformation need only be carried out to that accuracy.

Now, the Tolman-Lemaître co-ordinate system is co-moving, meaning that the time axis is aligned with the cosmological fluid's four velocity. The conformal Newtonian co-ordinate system is not co-moving, so at each event in space-time the two sets of co-ordinate axes must be related to one another by a radial boost transformation determined by the fluid four velocity (as well as, possibly, a rescaling of the co-ordinates). This is illustrated in Figure 5.1.

We know the contravariant components of the fluid four velocity in both co-ordinate systems so it is useful to write out its transformation equation. In the Tolman case the fluid velocity components are trivial since the co-ordinate system is co-moving and the t_{T} co-ordinate marks out intervals of local proper time so

$$u^{\alpha_{\mathrm{T}}} = \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix}. \tag{5.1.56}$$

In the Newtonian case, the fluid velocity components can be found in (C.9.27), and are

$$u^{\alpha} = \begin{pmatrix} 1 - \Phi & \delta u_r & 0 & 0 \end{pmatrix}. \tag{5.1.57}$$

From these, the co-ordinate transformation equation, $u^{\alpha} = \frac{\mathrm{d}x^{\alpha}}{\mathrm{d}x_{\mathrm{T}}{}^{\beta}} u^{\beta_{\mathrm{T}}}$, gives

$$u^{\alpha} = \frac{\mathrm{d}x^{\alpha}}{\mathrm{d}x_{\mathrm{T}}{}^{\beta}} u^{\beta_{\mathrm{T}}} = \frac{\mathrm{d}x^{\alpha}}{\mathrm{d}t_{\mathrm{T}}}$$

so

$$\frac{\mathrm{d}t}{\mathrm{d}t_{\mathrm{T}}} = 1 - \Phi, \qquad \frac{\mathrm{d}r}{\mathrm{d}t_{\mathrm{T}}} = \delta u_r. \tag{5.1.58}$$

The t_{T} unit vector has given us two of the transformation matrix components. We can use the r_{T} unit vector to get two more. It is fairly easy to obtain the contravariant components of the radial unit vector, $e_{r_{\mathrm{T}}}$, in the Tolman co-ordinate system, they are

$$e_{r_{\mathrm{T}}}{}^{\alpha_{\mathrm{T}}} = \begin{pmatrix} 0 & \frac{\sqrt{1 - k_{\mathrm{T}} r_{\mathrm{T}}{}^2}}{(a_{\mathrm{T}} r_{\mathrm{T}})_{,r_{\mathrm{T}}}} & 0 & 0 \end{pmatrix}. \tag{5.1.59}$$

In the Newtonian co-ordinate system, it can be easily verified that

$$e^{\alpha} = \begin{pmatrix} a\delta u_r & \frac{1}{a}(1 + \Phi) & 0 & 0 \end{pmatrix} \tag{5.1.60}$$

is an outwardly-directed vector for which $g_{\mu\nu} e^{\mu} e^{\nu} = 1$, and $g_{\mu\nu} u^{\mu} e^{\nu} = 0$. Being an

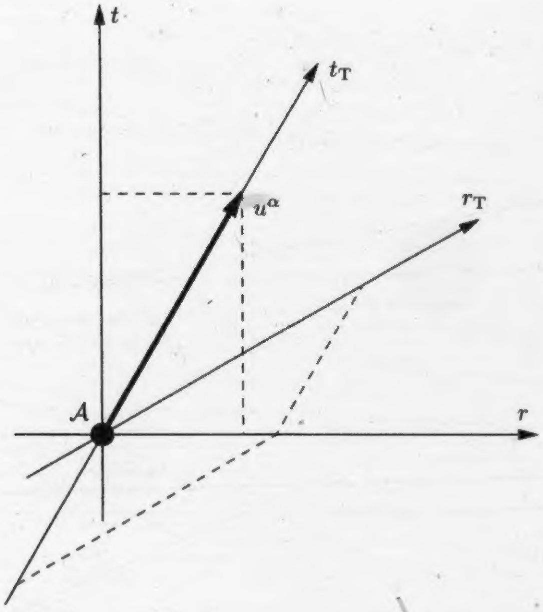


Figure 5.1: The schematic relationship between conformal Newtonian co-ordinates and Tolman-Lemaître co-ordinates at the space-time event $\mathcal{A}$. The diagram is drawn in the preferred reference frame of the Newtonian co-ordinate system; u^α is the four velocity of the perturbed cosmological fluid at $\mathcal{A}$; and the diagram shows how the Tolman time and radial co-ordinate directions are defined, relative to the Newtonian ones, by the fluid four velocity. For illustration, the dashed lines show how one unit of $\mathrm{d}t_\mathrm{T}$ is projected onto the t and r axes, and how one unit of $\mathrm{d}r$ is projected onto the t_T and r_T axes.

outward-directed space-like unit vector perpendicular to the fluid four-velocity, it must be the same vector as ${e_{r_\mathrm{T}}}^{\alpha_\mathrm{T}}$ above. The component transformation equation then leads to

$$e^\alpha = \frac{\mathrm{d}x^\alpha}{\mathrm{d}{x_\mathrm{T}}^\beta}{e_{r_\mathrm{T}}}^{\beta_\mathrm{T}} = \frac{\mathrm{d}x^\alpha}{\mathrm{d}r_\mathrm{T}}\frac{\sqrt{1-k_\mathrm{T}{r_\mathrm{T}}^2}}{(a_\mathrm{T}r_\mathrm{T})_{,r_\mathrm{T}}} \tag{5.1.61}$$

so

$$\frac{\mathrm{d}t}{\mathrm{d}r_\mathrm{T}} = \frac{(a_\mathrm{T}r_\mathrm{T})_{,r_\mathrm{T}}}{\sqrt{1-k_\mathrm{T}{r_\mathrm{T}}^2}}a\delta u_r \qquad \frac{\mathrm{d}r}{\mathrm{d}r_\mathrm{T}} = \frac{(a_\mathrm{T}r_\mathrm{T})_{,r_\mathrm{T}}}{\sqrt{1-k_\mathrm{T}{r_\mathrm{T}}^2}}\frac{1}{a}(1+\Phi). \tag{5.1.62}$$

Because of the degrees of freedom contained in a_T and k_T, on any *single* space-like hyper-surface we are free to choose any mapping of physical radii to values of r_T we please as long as the mapping is differentiable, one-to-one and onto. Let us choose to label events on our boundary hyper-surface with $r_\mathrm{T} = r$. This means that, on the boundary, we've chosen[8]

$$\frac{\mathrm{d}r}{\mathrm{d}r_\mathrm{T}} = 1. \tag{5.1.63}$$

This turns (5.1.62) into

$$\frac{\mathrm{d}t}{\mathrm{d}r_\mathrm{T}} = a^2\delta u_r \qquad (1-\Phi)a^2 = \frac{(a_\mathrm{T}r_\mathrm{T})_{,r_\mathrm{T}}}{\sqrt{1-k_\mathrm{T}{r_\mathrm{T}}^2}}. \tag{5.1.64}$$

The only other transformation matrix components we will need are the ones involving the angular co-ordinates. For now, it is sufficient to recognize that the radial boost transformation does not influence the two co-ordinates not shown in the Figure 5.1, θ and ϕ, so we can say that

$$\frac{\mathrm{d}\theta}{\mathrm{d}{x_\mathrm{T}}^\mu} \propto {\delta^\theta}_\mu, \tag{5.1.65}$$

$$\frac{\mathrm{d}\phi}{\mathrm{d}{x_\mathrm{T}}^\mu} \propto {\delta^\phi}_\mu. \tag{5.1.66}$$

This completes the inventory of co-ordinate transformation matrix components required for the analysis. We have, mostly, established the co-ordinate mapping across the boundary including the specification of one of the degrees of freedom: the relationship between the Tolman radial co-ordinate and the Newtonian radial co-ordinate. Let's move on to considering the actual intrinsic geometry of the boundary hyper-surface.

Using our knowledge of the transformation matrix components, we can expand the sums in the transformation equation for the metric tensors, to get

---

[8] The question of what $\mathrm{d}r/\mathrm{d}r_\mathrm{T}$ needs to be in order to maintain the desired relationship between r_T and r on the boundary hyper-surface is not this straight-forward. The statement $\mathrm{d}r/\mathrm{d}r_\mathrm{T} = 1$ means that the change in the Newtonian r co-ordinate as one moves 1 unit in the r_T direction must be 1. Below, we will choose our boundary hyper-surface to be one of constant t in the Newtonian co-ordinate system. Demanding that $r_\mathrm{T} = r$ on this surface is *not* consistent with the stated rate of change of the one with the other (refer to Figure 5.1). By adding a hypothetical first-order term to $\mathrm{d}r/\mathrm{d}r_\mathrm{T}$, however, one finds that the discrepancy introduced by our "inconsistency" is actually second-order. Because of this, we can get away with some sloppiness here.

$$g_{\mathrm{T}t_\mathrm{T}t_\mathrm{T}} = \left(\frac{\mathrm{d}t}{\mathrm{d}t_\mathrm{T}}\right)^2 g_{tt} + \left(\frac{\mathrm{d}r}{\mathrm{d}t_\mathrm{T}}\right)^2 g_{rr}, \tag{5.1.67a}$$

$$g_{\mathrm{T}r_\mathrm{T}r_\mathrm{T}} = \left(\frac{\mathrm{d}t}{\mathrm{d}r_\mathrm{T}}\right)^2 g_{tt} + \left(\frac{\mathrm{d}r}{\mathrm{d}r_\mathrm{T}}\right)^2 g_{rr}, \tag{5.1.67b}$$

$$g_{\mathrm{T}\theta_\mathrm{T}\theta_\mathrm{T}} = \left(\frac{\mathrm{d}\theta}{\mathrm{d}\theta_\mathrm{T}}\right)^2 g_{\theta\theta}, \tag{5.1.67c}$$

$$g_{\mathrm{T}\phi_\mathrm{T}\phi_\mathrm{T}} = \left(\frac{\mathrm{d}\phi}{\mathrm{d}\phi_\mathrm{T}}\right)^2 g_{\phi\phi}. \tag{5.1.67d}$$

(5.1.67c) tells us that

$$\left(\frac{\mathrm{d}\theta}{\mathrm{d}\theta_\mathrm{T}}\right)^2 = \frac{{a_\mathrm{T}}^2(t_\mathrm{T}, r_\mathrm{T}){r_\mathrm{T}}^2}{[1 - 2\Phi(t,r)]a^2(t)r^2}$$

but the right-hand side does not depend on the angular co-ordinates and since θ needs to be $\in [0, \pi]$ in both co-ordinate systems, we must have $\frac{\mathrm{d}\theta}{\mathrm{d}\theta_\mathrm{T}} = 1$ and $\theta_\mathrm{T} = \theta$ and, therefore,

$${a_\mathrm{T}}^2(t_\mathrm{T}, r_\mathrm{T}){r_\mathrm{T}}^2 = [1 - 2\Phi(t,r)]a^2(t)r^2.$$

Similarly, knowing now that $\theta_\mathrm{T} = \theta$ and that ϕ must be $\in [0, 2\pi]$ in both co-ordinate systems, (5.1.67d) gives us this result again and that $\phi_\mathrm{T} = \phi$. Using our choice of $r_\mathrm{T} = r$, the result above simplifies to

$$a_\mathrm{T}(t_\mathrm{T}, r) = [1 - \Phi(t,r)]\, a(t) \tag{5.1.68}$$

on the boundary hyper-surface. Using this expression for a_T on the boundary hyper-surface in (5.1.64) and solving for k_T gives (note: $\frac{\mathrm{d}}{\mathrm{d}r_\mathrm{T}} = \frac{\mathrm{d}r}{\mathrm{d}r_\mathrm{T}}\frac{\mathrm{d}}{\mathrm{d}r} + \frac{\mathrm{d}t}{\mathrm{d}r_\mathrm{T}}\frac{\mathrm{d}}{\mathrm{d}t}$)

$$k_\mathrm{T}(r) = \frac{2\Phi_{,r}}{r} - \frac{2Ha^2\delta u_r}{r}. \tag{5.1.69}$$

Using all of the information obtained so far, (5.1.67a) gives, simply, $-1 = -1$ so (5.1.67a) is redundant and we have run out of constraints.

Our inventory so far: we have mappings for all three spatial co-ordinates on the boundary hyper-surface, all the co-ordinate transformation matrix components needed to convert from Tolman to Newtonian co-ordinates, as well as rules for computing a_T and k_T at any event on the boundary hyper-surface. There is only one piece missing and that is a rule for computing $\frac{\mathrm{d}}{\mathrm{d}t_\mathrm{T}}a_\mathrm{T}$ on the hyper-surface. Note that we have yet to specify the boundary hyper-surface. All of the above holds regardless of the actual hyper-surface chosen to be the interface. By now choosing a specific boundary, we can enforce the equality of the hyper-surface's extrinsic curvature in each co-ordinate system. This will provide the additional constraints needed to specify the remainder of the mapping.

Let's choose to stitch the two descriptions of geometry together along a surface of constant t in the conformal Newtonian co-ordinate system. The extrinsic curvature of a space-like hyper-surface is the second-rank tensor given by the covariant derivative of the unit surface normal vector with respect to each of the three directions in the surface. Using the sign convention of [65], the extrinsic curvature tensor is

$$K_{ij} = -n_{i\,;\,j} = -n_{i\,,\,j} + \Gamma^{\alpha}{}_{ij} n_{\alpha} \tag{5.1.70}$$

where n^{α} is the unit vector normal to the surface of interest and i and j enumerate the three spatial directions parallel to the hyper-surface. Working in the Newtonian co-ordinate system, for a surface of constant t the components of the positive-time directed surface normal are

$$n^{\alpha} = \begin{pmatrix} 1-\Phi & 0 & 0 & 0 \end{pmatrix}, \qquad n_{\alpha} = \begin{pmatrix} 1+\Phi & 0 & 0 & 0 \end{pmatrix}. \tag{5.1.71}$$

We will only need $\Gamma^{0}{}_{ij}$ and, since the metric tensor is diagonal, only $\Gamma^{0}{}_{ii}$ will be non-zero and these are given by (C.1.5a),

$$\begin{aligned} \Gamma^{0}{}_{ii} &= -\frac{1}{2}\frac{1}{g_{tt}}\frac{\mathrm{d}}{\mathrm{d}t} g_{ii}. \\ &= a^2 \left[(1-4\Phi)\frac{\dot{a}}{a} - \dot{\Phi} \right] \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2\theta \end{pmatrix}. \end{aligned} \tag{5.1.72}$$

Using (5.1.71) and (5.1.72), the extrinsic curvature for a surface of constant t in the Newtonian space-time is

$$K_{ij} = a^2 \left[(1-3\Phi)\frac{\dot{a}}{a} - \dot{\Phi} \right] \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2\theta \end{pmatrix}. \tag{5.1.73}$$

We now need to compute the extrinsic curvature of the same surface in the Tolman co-ordinate system. This is how we'll accomplish the task: we'll transform the surface's normal vector to the Tolman co-ordinate system; compute its covariant derivative, $K_{\mathrm{T}\alpha_{\mathrm{T}}\beta_{\mathrm{T}}} = -n_{\alpha_{\mathrm{T}}\,;\,\beta_{\mathrm{T}}}$, in that co-ordinate system; transform the result back to the Newtonian co-ordinate system thereby projecting the Tolman derivatives onto the Newtonian spatial directions; finally, equate the spatial components of the resulting tensor, $K_{\mathrm{T}ij}$, to K_{ij} above in order to obtain an expression for $\frac{\mathrm{d}}{\mathrm{d}t_{\mathrm{T}}} a_{\mathrm{T}}$. Since there is only one piece of information we need to obtain from this procedure, it should be enough to equate just one non-trivial component of $K_{\mathrm{T}ij}$ to the corresponding component of K_{ij}. Recalling the co-ordinate transformation matrices obtained previously, the easiest non-trivial components to transform from the Tolman co-ordinate system back to the Newtonian co-ordinate system are $K_{\mathrm{T}\theta\theta} = K_{\mathrm{T}\theta_{\mathrm{T}}\theta_{\mathrm{T}}}$ and $K_{\mathrm{T}\phi\phi} = K_{\mathrm{T}\phi_{\mathrm{T}}\phi_{\mathrm{T}}}$, so we'll aim for computing just one of these rather than the entire tensor.

Transforming the unit normal vector in (5.1.71) to the Tolman co-ordinate system gives

$$\begin{aligned} n_{\alpha_{\mathrm{T}}} &= \frac{\mathrm{d}x^{\beta}}{\mathrm{d}x_{\mathrm{T}}{}^{\alpha}} n_{\beta} \\ &= \begin{pmatrix} \frac{\mathrm{d}t}{\mathrm{d}t_{\mathrm{T}}}(1+\Phi) & \frac{\mathrm{d}t}{\mathrm{d}r_{\mathrm{T}}}(1+\Phi) & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 & a^2\delta u_r & 0 & 0 \end{pmatrix}. \end{aligned} \tag{5.1.74}$$

Like the Newtonian metric tensor, the Tolman metric tensor is also diagonal so we can once again make use of (C.1.5a) but now we need many more components than in the Newtonian case since the unit vector is more complex and, in principle, we must compute all the components of its covariant derivative (although we actually won't be). In principle, we need

$$\begin{aligned} {\Gamma^{t_{\mathrm{T}}}}_{i_{\mathrm{T}}i_{\mathrm{T}}} &= -\frac{1}{2}\frac{1}{g_{\mathrm{T}t_{\mathrm{T}}t_{\mathrm{T}}}}\frac{\mathrm{d}}{\mathrm{d}t_{\mathrm{T}}} g_{\mathrm{T}i_{\mathrm{T}}i_{\mathrm{T}}}, & {\Gamma^{r_{\mathrm{T}}}}_{i_{\mathrm{T}}i_{\mathrm{T}}} &= -\frac{1}{2}\frac{1}{g_{\mathrm{T}r_{\mathrm{T}}r_{\mathrm{T}}}}\frac{\mathrm{d}}{\mathrm{d}r_{\mathrm{T}}} g_{\mathrm{T}i_{\mathrm{T}}i_{\mathrm{T}}}, \\ {\Gamma^{r_{\mathrm{T}}}}_{r_{\mathrm{T}}r_{\mathrm{T}}} &= \frac{1}{2}\frac{1}{g_{\mathrm{T}r_{\mathrm{T}}r_{\mathrm{T}}}}\frac{\mathrm{d}}{\mathrm{d}r_{\mathrm{T}}} g_{\mathrm{T}r_{\mathrm{T}}r_{\mathrm{T}}}, & {\Gamma^{r_{\mathrm{T}}}}_{t_{\mathrm{T}}r_{\mathrm{T}}} &= \frac{1}{2}\frac{1}{g_{\mathrm{T}r_{\mathrm{T}}r_{\mathrm{T}}}}\frac{\mathrm{d}}{\mathrm{d}t_{\mathrm{T}}} g_{\mathrm{T}r_{\mathrm{T}}r_{\mathrm{T}}}. \end{aligned}$$

All other Christoffel symbols are either 0 or not needed on account of the components of the unit vector in (5.1.74). Keeping in mind the components of the unit vector, ${\Gamma^{r_{\mathrm{T}}}}_{\mu_{\mathrm{T}}\nu_{\mathrm{T}}}$ need only be found to 0th order in the perturbations. Also, remembering that our only goal at this point is to determine $\frac{\mathrm{d}}{\mathrm{d}t_{\mathrm{T}}}a_{\mathrm{T}}$ in terms of things known in the Newtonian co-ordinate system, we are free to simplify things by substituting the known expressions for the Tolman metric components in terms of the Newtonian ones as long as we leave all $\frac{\mathrm{d}}{\mathrm{d}t_{\mathrm{T}}}a_{\mathrm{T}}$'s alone.

Examining the Christoffel symbols shows that the easiest to compute of the two angular-angular covariant derivatives of $n_{\alpha_{\mathrm{T}}}$ is

$$\begin{aligned} K_{\mathrm{T}\theta\theta} = K_{\mathrm{T}\theta_{\mathrm{T}}\theta_{\mathrm{T}}} &= -n_{\theta_{\mathrm{T}};\theta_{\mathrm{T}}} \\ &= -n_{\theta_{\mathrm{T}},\theta_{\mathrm{T}}} + {\Gamma^{t_{\mathrm{T}}}}_{\theta_{\mathrm{T}}\theta_{\mathrm{T}}} n_{t_{\mathrm{T}}} + {\Gamma^{r_{\mathrm{T}}}}_{\theta_{\mathrm{T}}\theta_{\mathrm{T}}} n_{r_{\mathrm{T}}} \\ &= -\frac{1}{2}\frac{1}{-1}\left(\frac{\mathrm{d}}{\mathrm{d}t_{\mathrm{T}}}({a_{\mathrm{T}}}^2{r_{\mathrm{T}}}^2)\right)(1) - \\ &\qquad \frac{1}{2}\frac{1}{a^2}\left(\left[\frac{\mathrm{d}}{\mathrm{d}r} + a^2\delta u_r \frac{\mathrm{d}}{\mathrm{d}t}\right]\left[(1-2\Phi)a^2r^2\right]\right)(a^2\delta u_r) \\ &= r^2(1-\Phi)a\frac{\mathrm{d}a_{\mathrm{T}}}{\mathrm{d}t_{\mathrm{T}}} - \frac{1}{2}\left[2(1-2\Phi)a^2r - 2\Phi_{,r}a^2r^2 + 2a^2\delta u_r a\dot{a}r^2\right]\delta u_r. \\ &= r^2(1-\Phi)a\frac{\mathrm{d}a_{\mathrm{T}}}{\mathrm{d}t_{\mathrm{T}}} - a^2r\delta u_r. \end{aligned}$$

Equating this to $K_{\theta\theta}$ we have

$$r^2(1-\Phi)a\frac{\mathrm{d}a_{\mathrm{T}}}{\mathrm{d}t_{\mathrm{T}}} - a^2r\delta u_r = a^2r^2\left[(1-3\Phi)\frac{\dot{a}}{a} - \dot{\Phi}\right]$$

or, in other words, on the boundary hyper-surface

$$\frac{\mathrm{d}a_{\mathrm{T}}}{\mathrm{d}t_{\mathrm{T}}} = \dot{a}\left[1 - 2\Phi - \frac{1}{H}\dot{\Phi} + \frac{\delta u_r}{Hr}\right]. \tag{5.1.75}$$

Finally, applying the geometry matching conditions in (5.1.68), (5.1.69) and (5.1.75) to the general expression for the collapse time in (5.1.27) gives us our alternate approximation for the collapse time, namely

$$\begin{aligned} t_{\mathrm{collapse}}(r) &\approx \pi\frac{\dot{a}_{\mathrm{T}}^2}{k_{\mathrm{T}}}\sqrt{\frac{a_{\mathrm{T}}{}^2}{k_{\mathrm{T}}}} \\ &= \pi\dot{a}^2\left[1 - 4\Phi - \frac{2}{H}\dot{\Phi} + \frac{2\delta u_r}{Hr}\right] a(1-\Phi)\left(\frac{2\Phi_{,r}}{r} - \frac{2Ha^2\delta u_r}{r}\right)^{-3/2} \\ &= \frac{\pi}{\sqrt{2}^3}H^{-1}\sqrt{raH}^3\left(\frac{\Phi_{,r}}{aH} - a\delta u_r\right)^{-3/2}. \end{aligned} \tag{5.1.76}$$

This result agrees exactly with the collapse time obtained using the previous technique in (5.1.54).

5.1.5 Approximating the Velocity Perturbations

An excellent discussion of the known properties of the large-scale velocity field of the cosmos can be found in [72, Chapter 7] and all of the numerical data that follows is taken from there. The velocity, measured with respect to the rest frame of the cosmic microwave background, of the material in a $50h^{-1}$ Mpc diameter region centred on us is $|a\delta u_i| \sim 2 \times 10^{-3}$. If this is a typical departure for the velocity field then the mean square is $\left\langle |a\delta u_i|^2 \right\rangle \sim 10^{-6}$. The radius of this departure is about 2×10^{-2} Hubble radii[9] and it can be found[10] that this is comparable to the mean square of $\Phi'/(aH)$ at the same length scale. It should be pointed out that this velocity is at the noise-limit of the data: $50h^{-1}$ Mpc corresponds to a red-shift of about 5000 kms^{-1} while the peculiar velocity being quoted is about 500 kms^{-1}, or about 10% of the red-shift; meanwhile the scatter is galaxy velocity measurements tends to be at the 20% level. Nevertheless, we must assume that both velocity perturbations and metric perturbations will play comparable roles in determining the statistics of collapse times.

It can be shown that in a flat, pressureless (dust-filled), universe, the velocity perturbation field quickly becomes highly correlated with the metric perturbation and so the quantity $\Phi'/(aH) - a\delta u_r$ really only has one degree of freedom. In particular, from the equations of motion for the perturbations we find that[11]

$$a\delta u_r \approx -\frac{2}{3}\frac{\Phi_{,r}}{aH}. \tag{5.1.77}$$

---

[9] See Table B.4.

[10] See Section 5.2.1 below.

[11] See (C.9.44).

This approximation reduces the collapse time to

$$t_{\text{collapse}}(r) = \pi\sqrt{\frac{27}{1000}}H^{-1}\sqrt{raH}^{3}\left(\frac{\Phi_{,r}}{aH}\right)^{-3/2}. \tag{5.1.78}$$

From a statistical stand-point, the result in (5.1.78) tells us that the properties of t_{collapse} are entirely determined by those of the single random variable $\frac{\Phi'}{aH}$. The first step in analyzing the statistics of the collapse time is to determine the properties of this random variable.

5.1.6 Evaluating t_{collapse}

The evaluation of the collapse time in (5.1.78) for a shell at a given radius requires knowledge of two parameters: the Hubble constant, H, and the rate of change in the radial direction of the Newtonian perturbation, Φ, per Hubble radius. At any fixed proper radius from the centre of the spherical distribution of matter, it must be expected that these two parameters evolve in time. In other words, it should be expected that the time it takes a shell to collapse is a function not only of the radius of the shell but also from *when* it is one measures its collapse time. In principle, knowledge of the geometry of space-time on any space-like hyper-surface allows one to evolve the geometry forward or backward to any other space-like hyper-surface, and so knowledge of Φ and its derivatives at one time can be used to specify these quantities at any other time. However, since the determination of a collapse time *is* an analysis of structure evolution, it is a duplication of effort to first evolve structures from one time to another in order to determine a collapse time referenced to the latter. It is, in other words, less of a contrivance to set the time from which we determine collapse times to be the time at which we have our knowledge of Φ.

At present, our most precise knowledge of the large-scale structure of Φ comes from measurements of the structure of the cosmic microwave background radiation. Although processes that have occurred in the universe since the time of the decoupling of matter from radiation do influence observations of the structure of the CMBR, such observations essentially amount to measurements of the structure of the universe at the time of the decoupling of matter from radiation.

To be definite, then, we will choose to measure the collapse times from the time of the decoupling of matter from radiation in our universe. Because the analysis that follows, of the statistical properties of the collapse time, holds regardless of the time from which the collapse times are referenced, the subscript "dec" shall be left off of the Hubble constant, H. However, whenever a numerical result is determined that requires a particular choice of H, the value in (5.1.2) shall be the value of the Hubble constant used.

5.2 The Collapse-Time Distribution

5.2.1 The Distribution of $\frac{1}{aH}\Phi'(r)$

If we can obtain the distribution for the quantity Φ' then (5.1.78) will allow us to obtain a distribution for the collapse times. The first step is to extract the spherically symmetric component of the metric perturbation from the results of the Newtonian analysis. Taking the plane-wave mode decomposition of the metric perturbation to be

$$\Phi(t,\vec{x}) = \frac{1}{\sqrt{2\pi}^3}\int \Phi_{\vec{k}}(t)\mathrm{e}^{\mathrm{i}\vec{k}\cdot\vec{x}}\,\mathrm{d}^3k \tag{5.2.1}$$

and introducing a spherical mode decomposition of the phase factor,[12]

$$\mathrm{e}^{\mathrm{i}\vec{k}\cdot\vec{x}} = \sum_{l=0}^{\infty}\mathrm{i}^l\sqrt{4\pi(2l+1)}\,\mathrm{j}_l(kr)Y_{l,0}(\alpha), \tag{5.2.2}$$

where α is the angle between $\vec{k}$ and $\vec{x}$ i.e. $\cos\alpha = \frac{1}{kx}\vec{k}\cdot\vec{x}$, gives

$$\Phi(t,\vec{x}) = \frac{1}{\sqrt{2\pi}^3}\int \Phi_{\vec{k}}(t)\left[\sum_{l=0}^{\infty}\mathrm{i}^l\sqrt{4\pi(2l+1)}\,\mathrm{j}_l(kr)Y_{l,0}(\alpha)\right]\mathrm{d}^3k.$$

Keeping only the $l = 0$ component of the spherical wave expansion leaves

$$\begin{aligned}\Phi(t,r) &= \frac{1}{\sqrt{2\pi}^3}\int \Phi_{\vec{k}}(t)\left[\sqrt{4\pi}\,\mathrm{j}_0(kr)\frac{1}{\sqrt{4\pi}}\right]\mathrm{d}^3k \\ &= \frac{1}{\sqrt{2\pi}^3}\int \mathrm{j}_0(kr)\Phi_{\vec{k}}(t)\,\mathrm{d}^3k \end{aligned}\tag{5.2.3}$$

where the spherical Bessel function of order 0 is [40, equation (9.87)]

$$\mathrm{j}_0(x) = \frac{\sin x}{x}. \tag{5.2.4}$$

An expression for the mode decomposition of the radial derivative of Φ can be found from (5.2.3) and one of the recurrence relations for the spherical Bessel functions, namely[13]

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[x^{-n}\,\mathrm{j}_n(x)\right] = -x^{-n}\,\mathrm{j}_{n+1}(x). \tag{5.2.5}$$

This recurrence relation gives us

$$\frac{\mathrm{d}}{\mathrm{d}r}\mathrm{j}_0(kr) = -k\,\mathrm{j}_1(kr) \tag{5.2.6}$$

---

[12] See [40, equation (10.45)].

[13] See [3, equation (11.164)].

and the radial derivative is, therefore,

$$\Phi'(t,r) = -\frac{1}{\sqrt{2\pi}^3}\int k\,\mathrm{j}_1(kr)\Phi_{\vec{k}}(t)\,\mathrm{d}^3k \tag{5.2.7}$$

where the spherical Bessel function of order 1 is found in (3.4.22). It will be found to be more convenient to work with the rate of change of Φ per Hubble radius rather than per co-ordinate interval. This is given by

$$\frac{\mathrm{d}}{\mathrm{d}(raH)}\Phi(t,r) = \frac{1}{aH}\Phi'(t,r) = -\frac{1}{\sqrt{2\pi}^3(aH)}\int k\,\mathrm{j}_1(kr)\Phi_{\vec{k}}(t)\,\mathrm{d}^3k. \tag{5.2.8}$$

Now, Φ' is a Gaussian random variable whose mean we already know is 0 so all we need to do is determine its variance and we have fully characterized its distribution. Since the mean is zero, the variance is just the mean square which, starting from the mode decomposition for Φ' in (5.2.8), is

$$\begin{aligned}\left\langle\left(\frac{\Phi'(t,r)}{aH}\right)^2\right\rangle &= \frac{1}{(2\pi)^3(aH)^2}\int k_1k_2\,\mathrm{j}_1(k_1r)\,\mathrm{j}_1(k_2r)\left\langle\Phi_{k_1}\Phi^*_{k_2}\right\rangle\,\mathrm{d}^3k_1\,\mathrm{d}^3k_2\\ &= \frac{1}{4\pi(aH)^2}\int \mathrm{j}_1^2(kr)\mathcal{P}_\Phi(k)\frac{\mathrm{d}^3k}{k},\end{aligned}$$

where (3.2.7) has been used to express the two-point spectral correlation function in terms of the field's power spectrum. Performing the angular integration, we have

$$\left\langle\left(\frac{\Phi'(t,r)}{aH}\right)^2\right\rangle = \frac{1}{(aH)^2}\int k\,\mathrm{j}_1^2(kr)\mathcal{P}_\Phi(k)\,\mathrm{d}k. \tag{5.2.9}$$

A count of the k's in the integrand reveals a problem: there is one single factor appearing in the numerator and a second k comes from the measure; that means we need at least four k's appearing in the denominator at large k or this integral will not converge. Each Bessel function $\mathrm{j}_1(kr)$ goes as k^{-1} for large k so together they contribute two.[14] For an $n_{s\Phi} = 1$ spectrum, $\mathcal{P}_\Phi(k)$ contributes none which leaves us short two factors of k.

This problem is not unexpected. If the field contains insufficiently suppressed modes all the way up to infinite frequency then the variance of Φ' must be unbounded since the field becomes non-differentiable. This divergence is not of the type discussed in Section 3.4.2 for $n_s = 1$ fields where there the culprit was long wavelengths. Here long wavelengths are irrelevant since we are looking at the properties of the gradient of Φ which is less sensitive to them. One can see this by ignoring the power spectrum and examining only the Bessel functions in (5.2.9) (left-overs from the gradient operation) which go from contributing two k's to the denominator at high frequencies to contributing two k's to the numerator at low. Because of the large number of k's in the numerator, the integral converges at long wavelengths for power spectra as "red" as $n_{s\Phi} > -3$.

---

[14] For the asymptotic behaviour of the Bessel function, see (3.4.23) and (3.4.24).

It is unfortunate that short wavelength modes are the dominant contribution to the variance since the model from which our spectrum has been obtained does not accurately represent — or even represent at all — the dynamics of such small structures. We are forced to filter short wavelength structure from the field's spectrum. This will be accomplished using the techniques explained in Sections 3.4.1 and 3.4.2. Introducing a short wavelength filter, $W_{k;r_0}$, of characteristic length r_0 and using (3.4.18) for the modified power spectrum one finds that the variance of $\frac{1}{aH}\Phi'$ from (5.2.9) becomes

$$\left\langle \left(\frac{\Phi'}{aH}\right)^2 \right\rangle = \frac{(2\pi)^3}{(aH)^2} \int k\, \mathrm{j}_1^2(kr) W_{k;r_0}^2 \mathcal{P}_\Phi(k)\, \mathrm{d}k. \qquad (5.2.10)$$

Let's now consider which window function we'll use. We need two additional factors of k in the denominator to make the integral converge at large k which means the window function needs to go at least as k^{-1} as $k \to \infty$. Both the top-hat window and the Gaussian window satisfy this condition and so either can be used. The top-hat window, however, goes just as k^{-2} at large k so it will only make the result converge for $-3 < n_{\mathrm{s}\Phi} < 3$. To relax this restriction, we will choose to use the Gaussian window which gives finite results for any $n_{\mathrm{s}\Phi} > -3$.

Introducing a power-law approximation for the power spectrum, it will be convenient to write it as

$$\mathcal{P}_\Phi(k) = \sigma^2 \left(\frac{k}{aH}\right)^{n_{\mathrm{s}\Phi}-1}. \qquad (5.2.11)$$

As many factors of $(aH)^{-1}$ have been added as there are factors of k. These factors are dimensionless and this notation ensures that the numerical value of the spectrum's amplitude, σ^2, is unaffected by redefinitions of the Robertson-Walker scale factor for fixed physical spectrum. The amplitude of the spectrum, σ^2, can be found by combining (4.4.20) with (4.5.1) which, for a $n_{\mathrm{s}\Phi} = 1$ field, gives

$$\sigma = \frac{3}{2}\delta_{\mathrm{H}} = 2.91 \times 10^{-5} \pm 7\% \qquad (5.2.12)$$

using the data in Appendix B. The power spectrum in (5.2.11) turns the variance of $\frac{1}{aH}\Phi'$ in (5.2.10) into

$$\left\langle \left(\frac{\Phi'}{aH}\right)^2 \right\rangle = \sigma^2 \left(\frac{1}{raH}\right)^{n_{\mathrm{s}\Phi}+1} \int x^{n_{\mathrm{s}\Phi}}\, \mathrm{j}_1^2(x) \exp\left[-\frac{r_0^2}{r^2}x^2\right] \mathrm{d}x \qquad (5.2.13)$$

where the Fourier transform of the Gaussian window in (3.4.27) has been used. It can be shown that this evaluates to

$$\left\langle \left(\frac{\Phi'}{aH}\right)^2 \right\rangle = \frac{\sigma^2}{4}\Gamma\Big(\frac{n_{\mathrm{s}\Phi}-1}{2}\Big) \left(\frac{1}{r_0 aH}\right)^{n_{\mathrm{s}\Phi}-1} (raH)^{-2} \times$$
$$\left\{ {}_1\mathrm{F}_1\Big(\frac{n_{\mathrm{s}\Phi}-1}{2}; \frac{1}{2}; -\Big(\frac{r}{r_0}\Big)^2\Big) + 1 + \right.$$

$$\frac{2}{3-n_{s\Phi}}\left(\frac{r_0}{r}\right)^2\left[{}_1\mathrm{F}_1\left(\frac{n_{s\Phi}-3}{2};-\frac{1}{2};-\left(\frac{r}{r_0}\right)^2\right)-1\right]\right\} \tag{5.2.14a}$$

for $n_{s\Phi} > -3$, $n_{s\Phi} \neq -1, 1, 3$, and

$$\left\langle\left(\frac{\Phi'}{aH}\right)^2\right\rangle = \frac{\sigma^2}{2}\left(\frac{1}{r_0aH}\right)^2\left\{{}_2\mathrm{F}_2\left(1,1;\frac{3}{2},3;-\left(\frac{r}{r_0}\right)^2\right) - {}_2\mathrm{F}_2\left(1,1;\frac{3}{2},2;-\left(\frac{r}{r_0}\right)^2\right)\right\} \tag{5.2.14b}$$

for $n_{s\Phi} = 1$. See Appendix C.12 for the derivation.

The smoothing procedure introduced a new issue that now needs settling and that is the choice of length scale, r_0, to smooth over. Before addressing this issue, it should be stated that in reality the physical fluctuation spectrum is not anticipated to be "flat" up to arbitrarily high frequencies. Physical processes in the early universe act to damp out fluctuations at small scales and the actual physical spectrum is described by the product of a power-law spectrum with a "transfer function." A useful summary of the transfer function associated with cold dark matter can be found in [56] where it is seen that the effect of the transfer function is to reduce $n_{s\Phi}$ by 2 at large k. The use of a simple window function in this document mimics this effect while remaining analytically tractable. As mentioned above, the top-hat window function in fact has exactly this effect — reducing the spectral index by 2 at large k — but the additional analytic simplicity of the Gaussian window causes it to be favoured over the top-hat window at this stage.

Regarding the window function's radius, there are essentially two choices:

1. Pick a fixed scale to smooth over, chosen so as to mimic the cold dark matter transfer function. This is achieved by setting $r_0 = \text{constant}$.
2. Always smooth over length scales comparable to, or at least determined by, the physical scale of interest. This is achieved by setting $r/r_0 = \text{constant}$.

A visual comparison of the two choices is illustrated in Figure 5.2 which shows the variance of $\frac{1}{aH}\Phi'(raH)$ for a perturbation field with $n_{s\Phi} = 1$ where the fixed-length smoothing radius has been set to $r_0aH = 1$. The vertical axis is normalized to the values of the curves at their intersection point which occurs at $raH = r_0aH = 1$ where the two procedures are identical. For the particular choices of parameters shown, the mean square at the smoothing radius is

$$\left\langle\left(\frac{\Phi'(raH=1)}{aH}\right)^2\right\rangle = 3.81 \times 10^{-2}\sigma^2 \tag{5.2.15}$$

and every factor of two reduction in the smoothing radius results in a factor of 4 increase in the mean square at that radius.

Apart from the location of their intersection point, the important feature to observe is that the fixed-scale smoothing procedure gives a variance that drops off several powers of r more slowly than the variable length procedure for radii greater than the smoothing radius. The fixed-scale choice also has a maximum that occurs just above the smoothing radius.

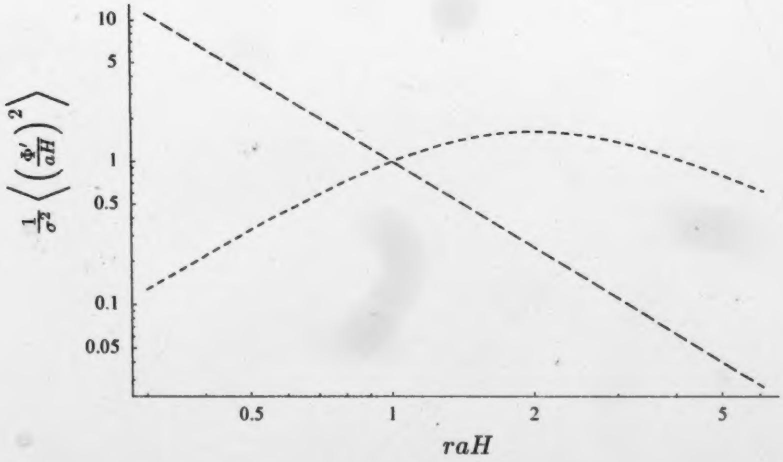


Figure 5.2: The mean square of $\frac{1}{aH}\Phi'(raH)$ after smoothing $n_{s\Phi} = 1$ perturbations using two techniques for choosing the smoothing scale. The short dashes show the behaviour for a fixed smoothing scale, $r_0aH = 1$, while the long dashes show the behaviour for a variable smoothing scale, $r_0/r = 1$. The vertical axis is normalized to the value of the two curves at their intersection point.

There are several points to be made with regard to which of the two procedures is "best." The fixed-scale smoothing procedure is, in a way, the more complex of the two since it requires the specification of a particular smoothing length while the sliding-scale procedure chooses this automatically. The biggest advantage of the fixed smoothing scale is that the statistics for all scales are being computed from the same fundamental perturbation field. A very important point regarding the sliding scale, though, can be made by returning to (5.2.13) and setting $r_0 = 0$ which recovers the result for the original, unfiltered, field. Doing so, and disregarding the divergent integral for the time-being, we see that the variance of Φ' then goes as $r^{-(n_{s\Phi}+1)}$ since the integral becomes independent of r. With a sliding scale given by $r/r_0 =$ constant, we obtain the exact same radial behaviour for the variance since the integral again becomes independent of r. Whether this is a point in favour or against a sliding scale is debatable. While the sliding scale does give the same radial behaviour as in the unfiltered case, the radial behaviour of the unfiltered field derives itself from a power spectrum that we know cannot be correct. Perhaps any procedure that reproduces that behaviour should be considered suspect.

In the computations to follow, the fixed-scale approach will be chosen and the scale set to the size of the smallest structures for which angular momentum is unlikely to play a role in their evolution. We choose this to be the Hubble radius,

$$r_0aH = 1. \tag{5.2.16}$$

Regarding the cold dark matter transfer function, the turn over to a lower spectral index is actually at shorter scales than this [22].

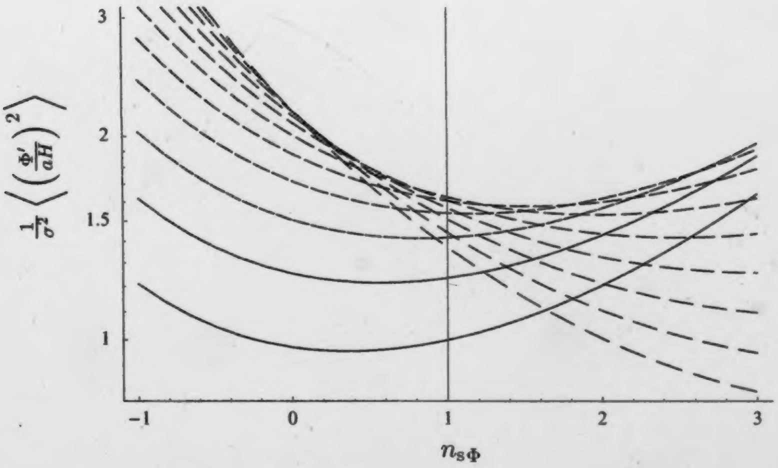


Figure 5.3: The mean square of $\frac{1}{aH}\Phi'(raH)$ as a function of $n_{s\Phi}$ for several radii with $r_0aH = 1$. The vertical axis is the same as in Figure 5.2. Each curve gives the behaviour at a different radius. The solid curve shows the behaviour at $r/r_0 = 1$; from there, in order of increasing gap length, the curves are for $r/r_0 = 1.2, 1.4, 1.6, \cdots, 2.8, 3.0$. The horizontal axis is marked at the spectral index illustrated in Figure 5.2.

It is worth investigating the behaviour of the variance as the spectral index is adjusted. Typical behaviour is shown in Figure 5.3 which is a plot of the variance of $\frac{1}{aH}\Phi'(raH)$ as a function of spectral index for a variety of radii using the same vertical axis as in Figure 5.2. To understand this plot, consider the set of curves as they cross $n_{s\Phi} = 1$. Together, they describe the behaviour of the variance as a function of r at that particular spectral index — in other words exactly what is shown in Figure 5.2. So, imagine that you are standing at the vertical axis in Figure 5.2 and looking down the horizontal axis over the hump of the curve; this is the pattern revealed in Figure 5.3 at $n_{s\Phi} = 1$ as one transitions from solid curve to increasing gap length. Visualizing this, it's seen that the envelope of the family of curves[15] in Figure 5.3 gives the amplitude of the peak in Figure 5.2 as a function of the spectral index. We can also see that the radius at which the peak occurs increases away from the smoothing radius as the spectral index drops, and decreases toward the smoothing radius as the spectral index increases.

5.2.2 The Distribution of $t_{\text{collapse}}(r)$

For the record, the distribution for $\frac{1}{aH}\Phi'$ is

$$P\left(\frac{\Phi'}{aH}\right) = \frac{1}{\sqrt{2\pi}\sqrt{\langle(\frac{\Phi'}{aH})^2\rangle}} \exp\left[-\frac{1}{2}\frac{(\Phi'/aH)^2}{\langle(\frac{\Phi'}{aH})^2\rangle}\right]. \tag{5.2.17}$$

---

[15] See [15, Section 4.3.1.6] for the definition and properties of the "envelope" of a family of curves.

Generally speaking, the mapping of the distribution for $\frac{1}{aH}\Phi'$ in (5.2.17) to a distribution for t_{collapse} is accomplished through the use of (D.1.1). In this case, however, due to the simple structure of the expression for the collapse time in (5.1.78), it is possible to use the special-case expression in (D.1.2) for quantities proportional to some power of a random variable. There are a number of complications to consider first, though. To begin with, although (5.2.17) does indeed provide the distribution for $\frac{1}{aH}\Phi'$, not all values of $\frac{1}{aH}\Phi'$ result in collapsing shells. The condition that a shell actually collapse was given in (5.1.55) and at this point examination of (5.1.78) reveals that the collapse condition translates into the condition

$$\frac{\Phi'}{aH} > 0. \tag{5.2.18}$$

If one applies this condition on $\frac{1}{aH}\Phi'$ to its distribution in (5.2.17) results in the distribution for $\frac{1}{aH}\Phi'$ *conditional on* the actual collapse of the shell in question,

$$P\left(\frac{\Phi'}{aH} \mid \text{collapse}\right) = \frac{2}{\sqrt{2\pi \left\langle (\frac{\Phi'}{aH})^2 \right\rangle}} \exp\left[-\frac{1}{2}\frac{(\Phi'/aH)^2}{\left\langle (\frac{\Phi'}{aH})^2 \right\rangle}\right], \quad \text{for } \frac{\Phi'}{aH} > 0, \tag{5.2.19}$$

where a factor of two has been included to adjust the distribution's normalization. The observation one should make at this point is that the mean of this distribution is no longer zero. All moments of this distribution can be found from[16]

$$\int_0^\infty x^n \exp(-ax^2)\,\mathrm{d}x = \frac{1}{2}a^{-\left(\frac{n+1}{2}\right)}\Gamma\left(\frac{n+1}{2}\right) \qquad \text{for } a > 0,\, n > -1. \tag{5.2.20}$$

From this, the n^{th} moment of the distribution in (5.2.19) is

$$\left\langle \left(\frac{\Phi'}{aH}\right)^n \mid \text{collapse} \right\rangle = \frac{2^{\frac{n}{2}}}{\sqrt{\pi}}\Gamma\left(\frac{n+1}{2}\right)\left\langle \left(\frac{\Phi'}{aH}\right)^2 \right\rangle^{\frac{n}{2}}. \tag{5.2.21}$$

I will carry the "conditional on collapse" modifier inside the average operations in order to distinguish averages of quantities conditional on the collapse of the shell from those of the same quantities when collapse is not a condition. With this in mind, the mean square of $\frac{1}{aH}\Phi'$ appearing on the right-hand side of this expression is the variance of the original, Gaussian, distribution. Since no quantum mechanical algebra is being used in this part of the document, there should be no danger of confusing this with "bra-ket" notation. Taking a look, specifically, at the mean of this new distribution we find that

$$\left\langle \frac{\Phi'}{aH} \mid \text{collapse} \right\rangle = \sqrt{\frac{2}{\pi}}\sqrt{\left\langle \left(\frac{\Phi'}{aH}\right)^2 \right\rangle}. \tag{5.2.22}$$

It is important to understand, however, that the mean collapse time is not finite despite the fact that $\frac{1}{aH}\Phi'$ has a non-zero mean when collapse is made a condition: although the integral giving the mean of $\frac{1}{aH}\Phi'$ is non-zero and converges, the integral giving the mean of $(\frac{1}{aH}\Phi')^{-\frac{3}{2}}$ does not converge and it is this power on which the collapse time depends.

---

[16] See [15, Section 1.1.3.4].

In principle, one could probably carry out the analysis to follow working entirely with the distribution for $\frac{1}{aH}\Phi'$ conditional on collapse but it will be found that this complicates things. It is easier to work with the full, Gaussian, distribution for $\frac{1}{aH}\Phi'$. All random variables derived from $\frac{1}{aH}\Phi'$ are, thus, allowed to take on values corresponding to $\frac{1}{aH}\Phi' \leq 0$ but with the understanding that the only interesting results are for values corresponding to $\frac{1}{aH}\Phi' > 0$.

Our second complication is the observation made above that even when collapse is made a condition, the expected collapse time is still infinite. Since the expected collapse time is infinite, it is helpful to study not the distribution of collapse times directly but the distribution for some map of collapse times that brings infinity to a finite value. Specifically, let's look at the distribution for $t_{\text{collapse}}^{-2/3}$. We'll choose this particular power of t_{collapse} not, primarily, because it maps infinity to 0 (which it does do) but because in choosing this, $\frac{1}{aH}\Phi'$ will appear linearly in the result. This greatly simplifies the analysis of the statistics. Rewriting (5.1.78) as an expression for $t_{\text{collapse}}^{-2/3}$ gives

$$t_{\text{collapse}}^{-2/3}(r) = \frac{10}{3\pi^{\frac{2}{3}}} H^{\frac{2}{3}} (raH)^{-1} \left(\frac{1}{aH}\Phi'\right). \tag{5.2.23}$$

Let's define a "reduced inverse collapse time" by factoring out some of the cosmological parameters:

$$t_{\text{collapse}}^{-2/3} = \frac{10}{3\pi^{\frac{2}{3}}} H^{\frac{2}{3}} \sigma \tau^{-\frac{2}{3}} \tag{5.2.24}$$

where

$$\tau^{-\frac{2}{3}}(raH) = \sigma^{-1}(raH)^{-1}\left(\frac{\Phi'(raH)}{aH}\right) \tag{5.2.25}$$

and where σ^2 is the amplitude of the power spectrum for Φ appearing in (5.2.11). We've taken out the numerical factors for the sake of brevity but left the r dependence along with enough a's and H's to allow us to write the radius as a multiple of the Hubble radius. The presence of σ in this expression has been chosen in such a way as to make the statistics of $\tau^{-\frac{2}{3}}$ independent of the normalization of Φ's power spectrum. $\tau^{-\frac{2}{3}}$ is dimensionless and the exponent reminds us which power of proper collapse time it is proportional to. The numerical relationship between $\tau^{-\frac{2}{3}}$ and t_{collapse} based on current cosmological data is shown in Figure 5.4.

Since $\tau^{-\frac{2}{3}}$ is linear in $\frac{1}{aH}\Phi'$, the mean square of the distribution for $\tau^{-\frac{2}{3}}$ can be easily found by using the mean square of $\frac{1}{aH}\Phi'$ from (5.2.14) directly in the square of the reduced inverse collapse time in (5.2.25) to get

$$\left\langle\left(\tau^{-\frac{2}{3}}\right)^2\right\rangle = \sigma^{-2}(raH)^{-2}\left\langle\left(\frac{\Phi'}{aH}\right)^2\right\rangle. \tag{5.2.26}$$

Just to reiterate: we are using the full Gaussian distribution for $\frac{1}{aH}\Phi'$. This means that

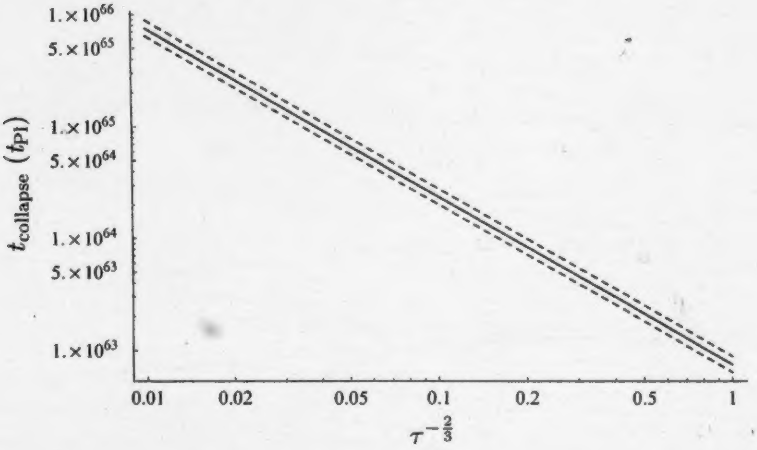


Figure 5.4: The relationship between t_{collapse} and $\tau^{-\frac{2}{3}}$. Here, $H = H_{\text{dec}} = 4.37 \times 10^{-57} \pm 6\%$ and $\sigma = 2.91 \times 10^{-5} \pm 7\%$ with the dashed lines indicating the error envelope. Note: $10^{60}\ t_{\text{Pl}} = 1.7$ Ga.

$\tau^{-\frac{2}{3}}$ is allowed to take on negative values. Of course, these correspond to imaginary collapse times, which are non-physical, so we must ignore them when looking at any specific behaviour. In the mean time, however, $\tau^{-\frac{2}{3}}$ can be considered to be just another random variable with a Gaussian distribution centred on 0, nothing more.

From the expression for the reduced inverse collapse time in (5.2.25) and the distribution map in (D.1.2), the distribution for $\tau^{-\frac{2}{3}}$ is

$$\begin{aligned} P(\tau^{-\frac{2}{3}}) &= \sigma(raH)P\left(\frac{\Phi'}{aH} = \sigma(raH)\tau^{-\frac{2}{3}}\right) \\ &= \frac{1}{\sqrt{2\pi}}\frac{\sigma(raH)}{\sqrt{\langle(\frac{\Phi'}{aH})^2\rangle}}\exp\left[-\frac{1}{2}\frac{\sigma^2(raH)^2}{\langle(\frac{\Phi'}{aH})^2\rangle}\left(\tau^{-\frac{2}{3}}\right)^2\right] \\ &= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\left\langle(\tau^{-\frac{2}{3}})^2\right\rangle}}\exp\left[-\frac{1}{2}\frac{1}{\left\langle(\tau^{-\frac{2}{3}})^2\right\rangle}\left(\tau^{-\frac{2}{3}}\right)^2\right] \end{aligned} \tag{5.2.27}$$

where the distribution for $\frac{1}{aH}\Phi'$ in (5.2.17) has been used. Of course we could also have just written this down immediately, knowing that $\tau^{-\frac{2}{3}}$ is Gaussian with the given mean and variance. We can integrate this distribution from a given $\tau^{-\frac{2}{3}} > 0$ to $+\infty$ in order to obtain the cumulative probability that a region has collapsed by the corresponding collapse time.

$$P\left(\text{collapse} \in [\tau^{-\frac{2}{3}}, \infty)\right) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\left\langle (\tau^{-\frac{2}{3}})^2 \right\rangle}} \int_{\tau^{-\frac{2}{3}}}^{\infty} \exp\left[-\frac{1}{2} \frac{1}{\left\langle (\tau^{-\frac{2}{3}})^2 \right\rangle} \left(\tau^{-\frac{2}{3}\prime}\right)^2 \right] \mathrm{d}\tau^{-\frac{2}{3}\prime}$$

$$= \frac{1}{2} \operatorname{erfc}\left(\frac{1}{\sqrt{2\left\langle (\tau^{-\frac{2}{3}})^2 \right\rangle}} \, \tau^{-\frac{2}{3}} \right) \tag{5.2.28}$$

where $\operatorname{erfc}(z)$ is the complementary error function defined as $\operatorname{erfc}(z) = 1 - \operatorname{erf}(z)$ and $\operatorname{erf}(z)$ is the error function defined as[17]

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z \exp(-t^2)\, \mathrm{d}t \tag{5.2.29}$$

and $\operatorname{erf}(\infty) = 1$.

The behaviour of the cumulative collapse probability, $P\left(\text{collapse} \in [\tau^{-\frac{2}{3}}, \infty)\right)$, in (5.2.28) is shown in Figure 5.5 and Figure 5.6. Figure 5.5 shows a family of cumulative probability curves parameterized by the size of the region, raH, with percentage contours marked. The particular curves indicated are at radii where the mean square of $\tau^{-\frac{2}{3}}$ has dropped to an integer fraction of the mean square at the smoothing radius. In other words, the curves are at the radii for which

$$\left\langle \left(\tau^{-\frac{2}{3}}(r)\right)^2 \right\rangle / \left\langle \left(\tau^{-\frac{2}{3}}(r_0)\right)^2 \right\rangle = 1, \frac{1}{2}, \frac{1}{3}, \cdots, \frac{1}{10}.$$

Figure 5.6 shows how the cumulative probability of collapse in $\tau^{-\frac{2}{3}}$ depends on the spectral index of the field, $n_{s\Phi}$. The plot is specifically the cumulative probability curve at the smoothing radius, that is at $raH = r_0 aH = 1$.

A useful metric for characterizing the curves in Figure 5.5 and Figure 5.6 is to identify the median collapse time — the time at which the probability that something that is going to collapse has collapsed is 1/2. Reducing each curve to a single characteristic number allows us to investigate how the collapse distributions depend on a wider variety of parameter choices. For convenience, let's introduce the number ζ such that

$$\zeta \equiv \operatorname{erf}^{-1}\left(\frac{1}{2}\right) = \operatorname{erfc}^{-1}\left(\frac{1}{2}\right) \approx 0.476936276\ldots, \tag{5.2.30}$$

and use this to write the median reduced inverse collapse time which, from (5.2.28), is

---

[17] Not to be confused with the Gaussian error integral defined as $\Phi(z) = (2\pi)^{-1/2} \int_{-\infty}^{z} \exp(-t^2/2)\, \mathrm{d}t$. See [15, Section 5.1.2.2.2].

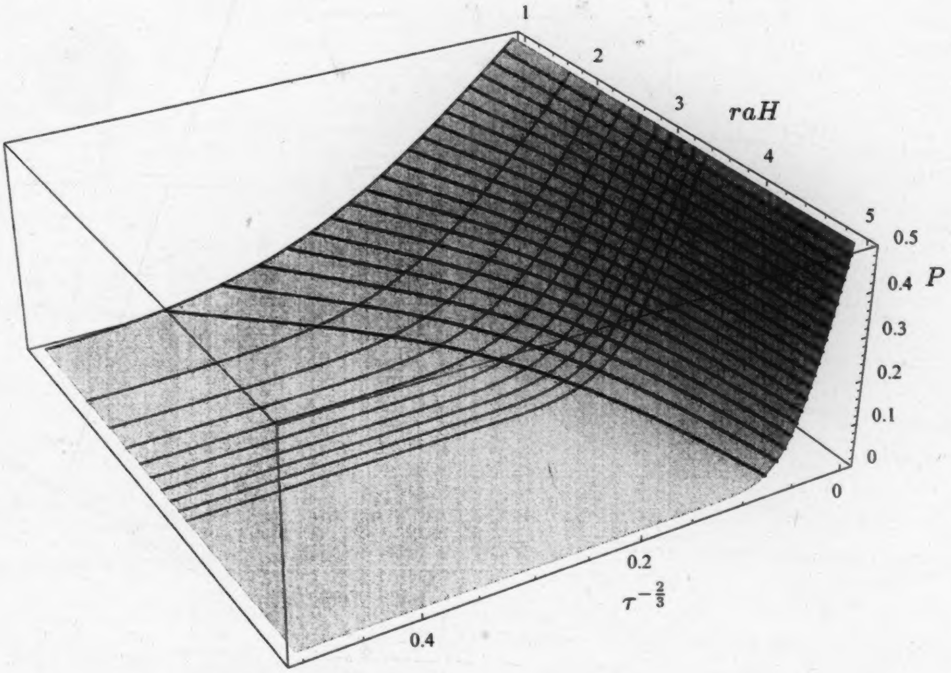


Figure 5.5: Cumulative probability of collapse $P\left(\text{collapse} \in [\tau^{-\frac{2}{3}}, \infty)\right)$ as a function of raH for a spectral index of $n_{s\Phi} = 1.01$ and smoothing radius of $r_0aH = 1$. The contours are drawn from 2.5% to 47.5% inclusively at 2.5% intervals. Also indicated are the particular cumulative probability curves for the first 10 radii at which the mean square of $\tau^{-\frac{2}{3}}$ is an integer fraction of the mean square at the smoothing radius.

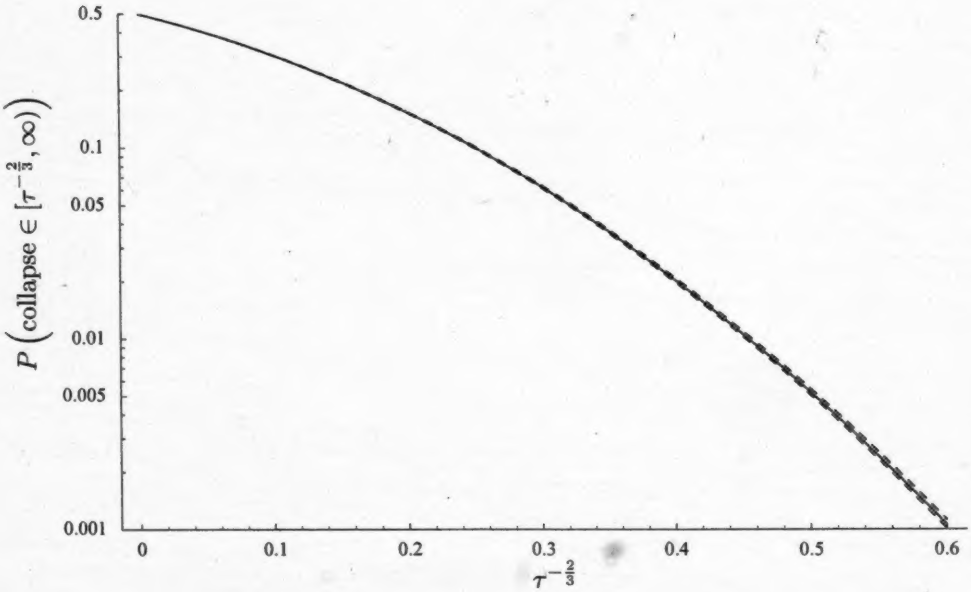


Figure 5.6: The cumulative probability of collapse, $P\left(\text{collapse} \in [\tau^{-\frac{2}{3}}, \infty)\right)$, at the smoothing radius, $raH = r_0aH = 1$, as it depends on the spectral index, $n_{s\Phi}$. The distribution is shown specifically for the expected value of $n_{s\Phi} = 1.01^{+0.09}_{-0.08}$ where the dashed lines indicate the envelope corresponding to the error range.

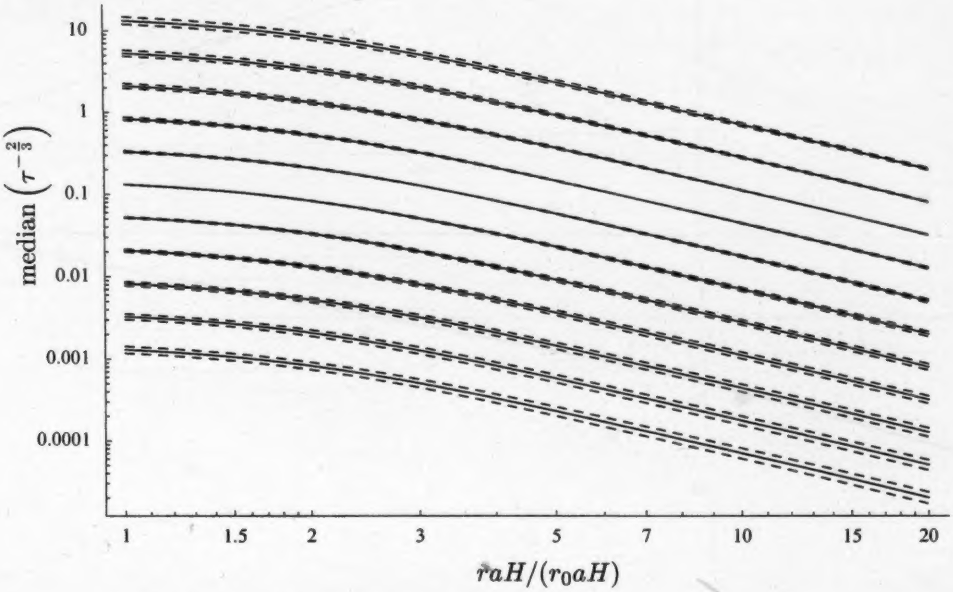


Figure 5.7: The median reduced inverse collapse time as a function of raH/r_0aH for several choices of smoothing radius and a range of spectral indices at each. The top-most solid curve corresponds to a smoothing radius of $r_0aH = 10^{-1}$, the bottom-most solid curve to $r_0aH = 10^1$ with the smoothing radii in-between occurring at multiples of $10^{0.2}$. All curves are drawn specifically for $n_{s\Phi} = 1.01^{+0.09}_{-0.08}$ with the dashed lines on either side of each solid curve showing the uncertainty envelope.

then

$$\text{median}\left(\tau^{-\frac{2}{3}}\right) = \zeta\sqrt{2\left\langle\left(\tau^{-\frac{2}{3}}\right)^2\right\rangle}. \qquad (5.2.31)$$

The behaviour of the median reduced inverse collapse time is shown in Figure 5.7.

5.2.3 Commentary

Throughout this computation, we have used a power spectrum for the metric perturbations consistent with what would be left over after a period of inflation involving a field with a $\frac{1}{2}m^2\phi^2$ potential and the particular structures being looked at have been taken to be larger than the Hubble radius.

For all radii and all spectral indices the cumulative probability of collapse goes to 0.5 as $\tau^{-\frac{2}{3}} \to 0$ (i.e. $t_{\text{collapse}} \to \infty$). Remember that the cosmological background has been assumed to be exactly critical. Gaussian fluctuations about this background thus lead to half of space being over-dense and the other half begin under-dense (even if only infinitesimally); hence the limit as $t_{\text{collapse}} \to \infty$ of half of space collapsing.

One can remark on how rapidly the distribution collapses to a δ-function at $t_{\text{collapse}} =$

∞. This shrinking to a δ-function is directly visible in Figure 5.5 and is more carefully quantified in Figure 5.7. From the large-r asymptotic behaviour of $\left\langle (\frac{\Phi'}{aH})^2 \right\rangle$ discussed in Appendix C.12 and the mean square of $\tau^{-\frac{2}{3}}$ in (5.2.26) we see that

$$\left\langle (\tau^{-\frac{2}{3}})^2 \right\rangle \sim \left(\frac{r}{r_0}\right)^{-6} \tag{5.2.32}$$

for $r \gg r_0$ and spectral indices near 1. One can understand the decrease of the inverse collapse-time's median with increasing radius as being a consequence of the fact that larger regions simply take longer to collapse so the median collapse time must increase.

Let me point out that there are only three parameters that determine the median reduced inverse collapse time: the spectral index n_s, the smoothing radius $r_0 aH$, and the radius of observation raH. Converting the reduced inverse collapse time to a proper collapse time requires, additionally, the spectral normalization σ and the Hubble constant H, but the reduced inverse collapse time itself is independent of these. With this in mind, a curious feature of the behaviour of its median is visible in Figure 5.7: it appears to be independent of n_s at median $\left(\tau^{-\frac{2}{3}}\right) \sim 0.1$, regardless of the choice of smoothing radius. That the median is insensitive to the spectral index somewhere is not surprising: the spectral index enters in via the variance of $\frac{1}{aH}\Phi'$ which, as a function of spectral index, has a minimum where $\mathrm{d}(\frac{1}{aH}\Phi')/\,\mathrm{d}n_s = 0$. These minima can be seen in Figure 5.3. It is curious that the result should be that the median inverse collapse time is insensitive to n_s always at median $\left(\tau^{-\frac{2}{3}}\right) \sim 0.1$, regardless of the choice of smoothing radius.

The choice of smoothing radius at this particular spectral index has a clear and straight-forward effect on the results: it simply shifts the over-all amplitude of the median reduced inverse collapse time without, apparently, changing its behaviour with respect to radius. Smaller smoothing radii increase the median which increases the likelihood of small proper collapse times. This is easily understood: smoothing at smaller scales is equivalent to "less" smoothing — the smoothed field looks more like the underlying field — so greater variation should be seen.

It must be pointed out that the computation performed in this section does not take into consideration the possibility that one might have made a measurement of part of the universe. This computation is only looking at the inherent variation in collapse times. We ourselves know, however, that our patch of the universe look very flat. The question we should ask, then, is: given that our visible patch of the universe appears to be close to critical now, how might our assessment of this change over time?

Before we answer this question, however, we first need to read some bad news. Throughout all of Section 5.1, in obtaining the collapse times in the Tolman-Lemaître approximation, it has been assumed that the material content of the universe is non-relativistic, pressureless, dust. Current observational evidence suggests that this is not the case; that in fact the universe is even at present undergoing a period of inflation.[18]

---

[18] See [10] for evidence for a cosmological constant from the CMB, [50] for a review and [71] for the details of evidence from baryon abundances and large-scale structure, and [79] for evidence from supernovae.

The parameters most consistent with current observations suggest that in the neighbourhood of 70% of the current energy density of the universe is in the form of an inflaton field. We really should investigate the consequence of this accelerating expansion on the collapse of structures in the universe. We will do this in the next section before returning to the question posed above.

5.3 Incorporating a Cosmological Constant

5.3.1 Dust Condensing out of an Inflating Background

In [95], Weinberg considered the ability of a dust fluctuation to collapse under its own gravity in the presence of an inflating background.[19] His purpose was to determine the constraints that needed to be placed on the rate of inflation or, more precisely, on the inflaton energy density given the anthropic requirement of our existence. The criterion he equated with our existence was that the cosmos needed to admit galaxy-sized gravitationally bound clumps of matter. His model involved a cosmological fluid consisting of two parts: an inflaton field with energy density $\rho_\Lambda = \Omega_\Lambda \rho_c$ and a dust field with mean energy density $\rho_b = \Omega \rho_c$. The inflaton field was taken to be homogeneous while the dust component was taken to have a single spherically symmetric fluctuation described by the contrast $\overline{\delta} = \delta\rho/\rho_b$.

Weinberg found that a region with a mean density contrast of $\overline{\delta}$ can collapse if the condition that[20]

$$\rho_\Lambda < \frac{500}{729} \frac{\delta\rho^3}{\rho_b^2} \tag{5.3.1}$$

is satisfied. Rewriting this in terms of the density parameters,

$$\Omega_\Lambda \rho_c < \frac{500}{729} \frac{\left(\Omega\rho_c\overline{\delta}\right)^3}{\Omega^2 {\rho_c}^2} \tag{5.3.2}$$

or, in other words, the region collapses if its mean density contrast satisfies

$$\overline{\delta}^3 > \frac{729}{500} \frac{\Omega_\Lambda}{\Omega}. \tag{5.3.3}$$

Using the values for our universe today from Appendix B (taking the dust component to be both the baryon and dark matter components of our universe) this becomes the condition

$$\overline{\delta} \geq 1.1339 \left(\frac{0.70h^2}{0.142}\right)^{\frac{1}{3}} \approx 1.52. \tag{5.3.4}$$

Let's apply this model to our fluctuations by computing the probability of finding a region of some size with a mean density contrast satisfying the condition in (5.3.4).

---

[19] This model has also been examined by Martel, Shapiro and Weinberg in [63] and is included in Weinberg's review in [97].

[20] See [95, equation (7)].

Of course our cosmological perturbations do not consist of a single isolated spherically symmetric ball but we will make the approximation that any neighbouring fluctuations will not have a significant effect on the gravitational binding of the region in question.

Integrating over the domain in (5.3.4) gives

$$P(\text{condensation}; r) = \int_{1.52}^{\infty} P\left(\overline{\delta}; r\right) \, \mathrm{d}\overline{\delta} \tag{5.3.5}$$

for the probability that a region of size r will condense out from the background inflation where $P(\overline{\delta}; r)$ is the distribution for $\overline{\delta}$ smoothed over radius r. Taking the underlying density perturbations to be a homogeneous isotropic Gaussian random field with 0 mean, $P\left(\overline{\delta}; r\right)$ is given by (3.4.31) which allows us to evaluate this integral by writing that

$$\begin{aligned} P(\text{condensation}; r) &= \frac{1}{\sqrt{2\pi\left\langle\overline{\delta}^2\right\rangle}} \int_{1.52}^{\infty} \exp\left[-\frac{1}{2}\frac{\overline{\delta}^2}{\left\langle\overline{\delta}^2\right\rangle}\right] \mathrm{d}\overline{\delta} \\ &= \frac{1}{2}\operatorname{erfc}\left(\frac{1.52}{\sqrt{2\left\langle\overline{\delta}^2\right\rangle}}\right). \end{aligned} \tag{5.3.6}$$

For a fluctuation spectrum, we will assume the curvature perturbations have $n_{\mathrm{s}\mathcal{R}} = 1$ and the horizon-entry contrast is normalized to the COBE data, in other words

$$\mathcal{P}_\delta(k) = \left(\frac{k}{aH}\right)^4 \delta_{\mathrm{H}}^2 \tag{5.3.7}$$

with δ_{H} given in Appendix B. This is a power spectrum with $n_{\mathrm{s}\delta} = 5$. Note that the assumption of $n_{\mathrm{s}\delta} = 5$, although consistent with observation, is not required in what follows. Being able to obtain analytic results only requires that $n_{\mathrm{s}\delta} > 1$. Since the mean of $\overline{\delta}$ is 0, its variance is simply its mean square which, from (3.4.32), is

$$\left\langle\overline{\delta}^2\right\rangle = (2\pi)^3 \int_{k=0}^{\infty} W_{k;r}^2 \mathcal{P}_\delta(k) \frac{\mathrm{d}k}{k}. \tag{5.3.8}$$

Using a Gaussian window this can be evaluated analytically for any $n_{\mathrm{s}\delta} > 1$. This result, from (3.4.35), is

$$\left\langle\overline{\delta}^2\right\rangle = \frac{\delta_{\mathrm{H}}^2}{(raH)^{n_{\mathrm{s}\delta}-1}} \frac{1}{2} \Gamma\left(\frac{n_{\mathrm{s}\delta}-1}{2}\right) \tag{5.3.9}$$

which now allows the computation of the condensation probability in (5.3.6). The behaviour of this probability as a function of region size for a range of spectral indices near $n_{\mathrm{s}\delta} = 5$ is shown in Figure 5.8. Note that the curves are not distributions: both r and $n_{\mathrm{s}\delta}$ are parameters for an absolute condensation probability. A detailed plot specifically for $n_{\mathrm{s}\delta} = 5.01^{0.09}_{0.08}$ is shown in Figure 5.9.

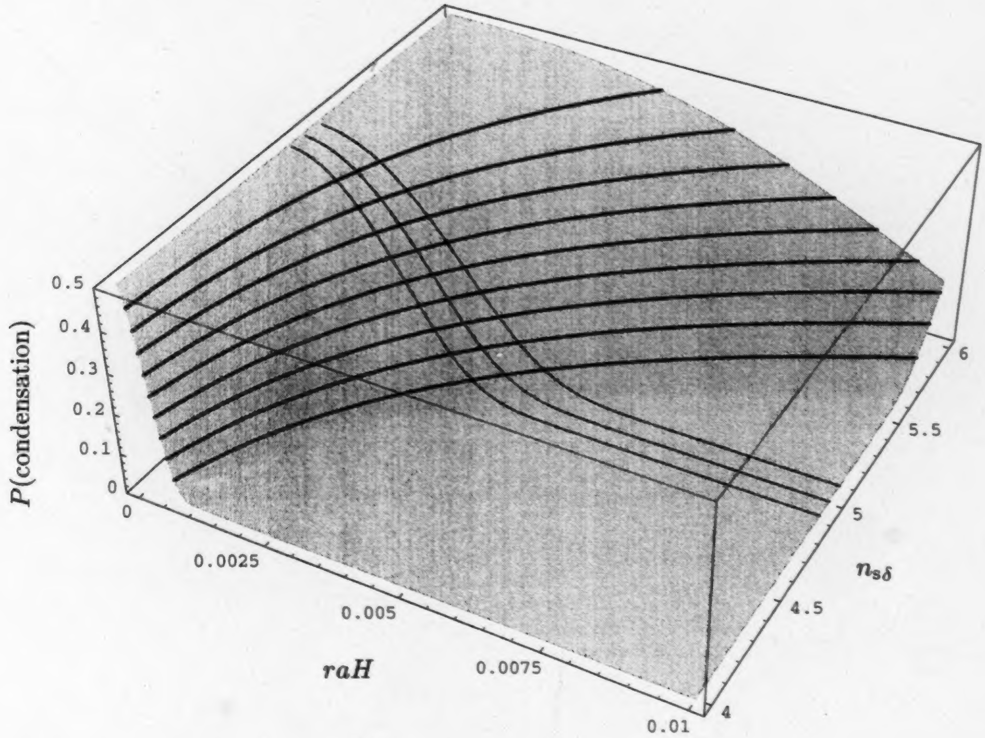


Figure 5.8: The probability of condensation as a function of region size for matter density spectral indices in the range $4 \leq n_{s\delta} \leq 6$, corresponding to $0 \leq n_{s\mathcal{R}} \leq 2$. The expected spectral index and its uncertainty envelope is illustrated at $n_{s\delta} = 5.01^{0.09}_{0.08}$ and the contours indicate probabilities from 5% through 45% inclusively in 5% increments.

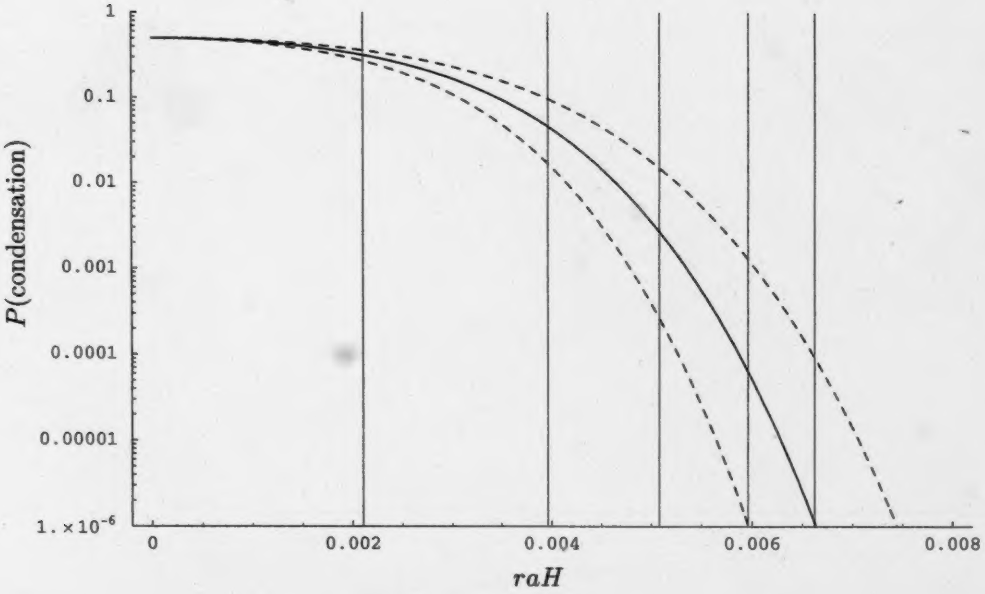


Figure 5.9: The probability of condensation as a function of region size for $n_{s\delta} = 5.01^{0.09}_{0.08}$. The dashed lines show the envelope for the indicated error range. The horizontal axis has been marked off, from left to right, with the radii corresponding to 1, 2, 3, 4, and 5 "σ" — unlikelihoods of collapse along the central curve corresponding to that many standard deviations.

5.3.2 Commentary

The horizontal axis in the plots presented above is raH which gives the radius of the region in question in units of the current Hubble radius. What can be seen from Figure 5.9 is that given our current knowledge of the universe's parameters, very little of anything with a radius larger than about 0.006× the Hubble radius, or about 26 Mpc, can be expected to be gravitationally bound. From the data in Appendix B, this is a diameter about 17 times that of a typical cluster of galaxies or just slightly larger than four times the diameter of the Virgo cluster.

A feature of the plots that should be noticed is that for all choices of spectral index displayed, the probability of condensation asymptotes to 0.5 as the radius of the window function shrinks to 0. Remember that the condition for gravitational condensation is that the mean density contrast within the region has to be greater than ~ 1.52. We have constructed the density contrast field to have a mean of 0 so one should expect 50% of the field to have a value greater than 0. How then have we ended up with 50% of the field having values larger than 1.52? The answer comes from remembering that random fields with spectral indices $n_s \geq 1$ have an unbounded variance on account of the power in short wavelength modes. Our field has a spectral index of $n_s = 5$. Smoothing the field, as done here to obtain the mean density contrast within a ball, filters out those high frequencies and gives a finite variance for the result but as the radius of the window function shrinks to zero we recover the properties of the underlying field. For a Gaussian random variable with 0 mean and an infinite variance, 50% of the samples will be larger than any finite positive number.

This brings up another issue, namely that our spectrum has been obtained assuming that linear perturbation theory is applicable and for that to be the case it's reasonable to consider it necessary for the density contrast to be small, $\delta \ll 1$. In this section we have done a computation that gives no results at all unless $\delta \gtrsim 1.52$ at least somewhere. There are a number of comments to make with regard to this. First of all, it should be noted that the result from Weinberg's paper in (5.3.1) was arrived at without relying on linear perturbation theory — it is exact in so far as the toy model it uses for the distribution of matter is correct (that of an isolated ideal top-hat pocket of dust). The only question here is that of the validity of our spectrum. For linear cosmological perturbation theory to be accurate, it is really only necessary for the perturbations to the geometry of spacetime, Φ, to be small. This is the case for even quite large energy density fluctuations as long as they are confined to small regions: recall that on scales shorter than the Hubble radius the density contrast is proportional to the Laplacian of the metric perturbation; see Appendix C.9.2. The other point to remember is that while the density contrast might be very large in places, its Fourier decomposition can still consist of only small-valued, linear, modes and this certainly is the case here. Recalling the power spectrum for the density contrast in (5.3.7) and the normalization in Appendix B we see that the the power spectrum only becomes larger than 1 for modes smaller than about 0.003× the Hubble radius. This agrees nicely with the data in Appendix B whence we see that the true scale at which modes have gone nonlinear in our observable universe is just a bit smaller than this at approximately 0.0027× the Hubble radius. Although this is squarely in the middle of the length scales of interest in the current results, it should not be considered to invalidate the cut-off — the result that little of anything larger than about 0.006× the Hubble radius will be gravitationally bound. We should, however, consider the small-scale asymptote of $P(\text{condensation}) = 0.5$ not to be representative of reality. On small scales the results do become sensitive to modes that really should have been treated non-linearly.[21]

The binding cut-off of 0.006× the Hubble radius, in a sense, gives us a short answer to the question posed in the previous section. Given that our visible universe appears to be critical, how does its apparent collapse time evolve over time? It doesn't: it will always be critical and its apparent collapse time will never deviate from being infinite — the universe will expand and cool and diffuse forever. This doesn't mean it's not worth working out an answer to our question anyway. We don't know what the "dark energy" driving the current period of inflation is. It is possible it could decay away, bringing the current period of inflation to an end and leaving a purely dust-filled universe in its wake. In that scenario, the answer to our question becomes interesting once more, although we can be less certain about the spectrum of fluctuations that will be seen at late times.

5.4 The Evolution of the Universe's Collapse-Time Distribution

In Section 5.2.2 we investigated the distribution of collapse times for post-inflationary patches of the universe. We'll now investigate how this distribution evolves over time given a single, instantaneous, observation of the collapse time. Of course, one doesn't observe a collapse time but rather infers it from measurements of cosmological parameters such as the mean density of the cosmological fluid and the Hubble constant. Any measurement of cosmological parameters can only be performed on that which is visible within some horizon. In principle the limit is the particle horizon but in practise some process in the universe's past will limit observations to a smaller volume. For example, the opacity of the plasma that was the cosmological fluid before matter-radiation decoupling prevents optical observations beyond a radius corresponding to the "surface of last scattering." As time elapses, such horizons expand outward from an observer's position allowing them access to a greater volume of the cosmos for their measurements. The question we are going to answer, then, is how might an observer's perception of the collapse time of the material surrounding them change as they are given knowledge of more of their surroundings.

Essentially, we need to compute a two-point joint distribution for the collapse time. Since the time dependence of the observer's perception of the universe's collapse time has its origin in the increasing volume of space available for measurement, the time dependence is ultimately determined by the *radial* two-point joint distribution.

As in Section 5.2.2, we will work with the reduced inverse collapse time and take its distribution to be the full, Gaussian, distribution. We need to introduce a second parametric radius. We will continue to use r_0aH as the smoothing radius and raH as the independent variable, and introduce r_1aH as the radius at which the observer's single observation of criticality has been made. In short, the statement of the problem is: the universe is smoothed at r_0aH, observed to be critical out to a radius of r_1aH and we ask for the collapse-time distribution at $raH > r_1aH$.

---

[21] A proper treatment of the question of what fraction of matter eventually ends up in collapsed objects can be found in [62].

There is the question of when, exactly, the observer will be able to see out to a distance raH. In fact, what we really want to know is how much proper time elapses between the observer being able to see out to r_1aH and the observer being able to see out to raH. At any point along his or her world line, an observer is able to see all of the space-time events coincident with their past light-cone. We are assuming, however, that some horizon prevents that light-cone from being extended beyond a proper radius of r_1a. The time along the observer's world line at which this is the case, we'll call t_1. At some later time, t, the light-cone can extend out to a proper radius of $ra > r_1a$. This arrangement is illustrated in Figure 5.10.[22] The question is: given r_1a, ra and $a(t)$, find $\Delta t = t - t_1$. We will do this to zeroth order in the perturbations.

From (2.1.13), $a(t) \propto t^{\frac{2}{3(1+w)}}$. Assuming dust ($w = 0$) and introducing a proportionality constant for bookkeeping, $a(t) = a_0t^{2/3}$. Now $H = \dot{a}/a = \frac{2}{3}t^{-1}$ so $t = \frac{2}{3}H^{-1}$ which we can use to set the time co-ordinate on the horizon. Combining these expressions, we can write $a(t)$ on the horizon as $a = a_0(\frac{2}{3}H^{-1})^{\frac{2}{3}}$. The past light-cones are null radial geodesics so from the Robertson-Walker line-element in (2.1.3), $\frac{1}{a(t)}\,\mathrm{d}t = \mathrm{d}r$. Integrating both sides of this gives, for any past light-cone,

$$\int_{\frac{2}{3}H^{-1}}^{t} \frac{1}{a(t')}\,\mathrm{d}t' = r$$

$$\frac{1}{a_0}\int_{\frac{2}{3}H^{-1}}^{t} t'^{-\frac{2}{3}}\,\mathrm{d}t' = r$$

$$\frac{3}{a_0}\left[t^{\frac{1}{3}} - \left(\frac{2}{3}H^{-1}\right)^{\frac{1}{3}}\right] = r.$$

Multiplying both sides by a on the horizon,

$$3\left(\frac{2}{3}H^{-1}\right)^{\frac{2}{3}}\left[t^{\frac{1}{3}} - \left(\frac{2}{3}H^{-1}\right)^{\frac{1}{3}}\right] = ra$$

$$3\left(\frac{2}{3}H^{-1}\right)^{\frac{2}{3}} t^{\frac{1}{3}} - 2H^{-1} = ra$$

$$27\left(\frac{2}{3}H^{-1}\right)^{2} t = \left(ra + 2H^{-1}\right)^{3}$$

$$Ht = \frac{1}{12}\left(raH + 2\right)^{3}.$$

---

[22] Remember that ra is a proper radius and raH is a proper radius expressed as a multiple of the Hubble radius. It has been convenient to blur the distinction between the two but now we need to be clear about it.

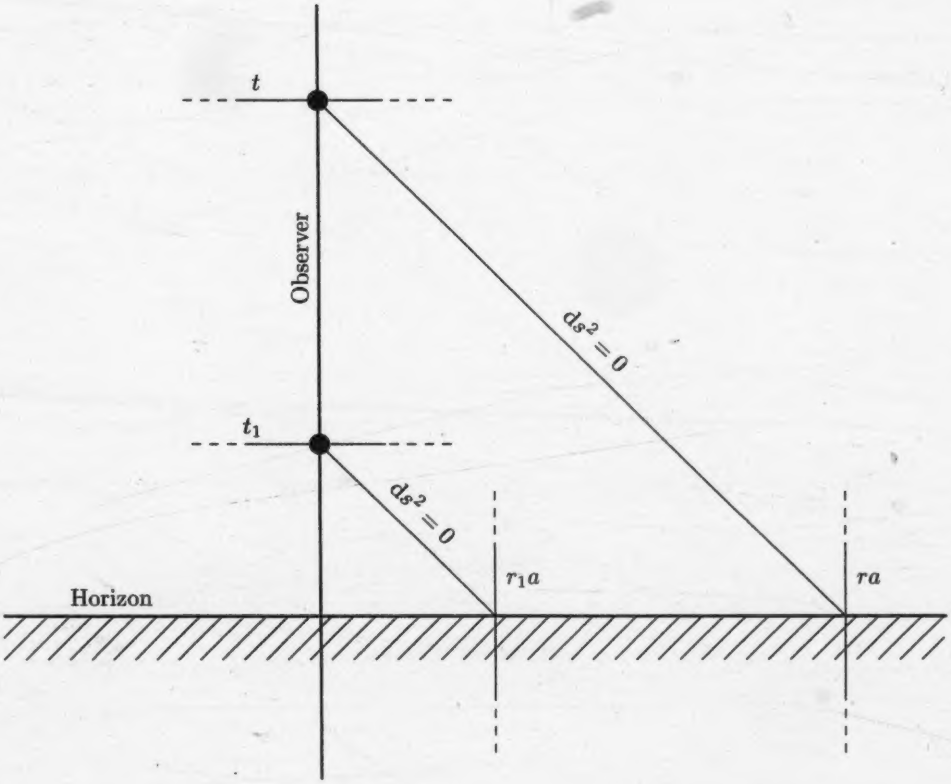

t
t_1
Observer
$ds^2 = 0$
$ds^2 = 0$
r_1a
ra
Horizon

Figure 5.10: The radius to which an observer can see as a function of time.

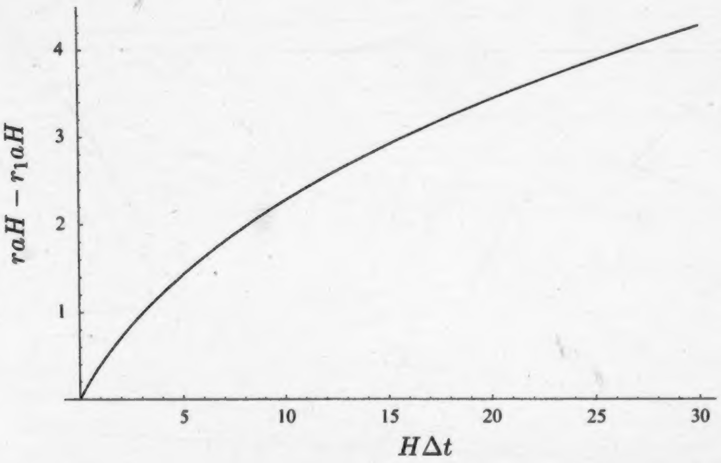


Figure 5.11: The increase of horizon radius as a function of elapsed time. For this plot, the horizon was set to an initial radius of $r_1aH = 1$.

Therefore, the time that elapses along the observer's world line between their being able to see out to r_1aH and their being able to see out to raH is

$$H\Delta t = \frac{1}{12}\left[(raH+2)^3 - (r_1aH+2)^3\right] \tag{5.4.1}$$

where $H\Delta t$ is the elapsed time in multiples of the Hubble time. This relationship is shown in Figure 5.11 for $r_1aH = 1$. It is very important to keep in mind that when we talk about radii in units of the Hubble radius and times in units of the Hubble time that these are specifically the Hubble radius and the Hubble time defined at the time of the horizon. In particular, they are *not* these dimensions as defined at the time of the observer's measurement at t_1. We can now proceed to compute the evolution of the collapse-time distribution.

5.4.1 The Radial Two-Point Joint Collapse-Time Distribution

Since $\tau^{-\frac{2}{3}}$ is a Gaussian random variable, its two-point joint distribution is given by (3.1.2) whence we see that

$$P\left(\tau^{-\frac{2}{3}}(r_1), \tau^{-\frac{2}{3}}(r)\right) = \frac{1}{2\pi\sqrt{\det\mathbb{B}}}\exp\left[-\frac{1}{2}\begin{pmatrix}\tau^{-\frac{2}{3}}(r_1) & \tau^{-\frac{2}{3}}(r)\end{pmatrix}\mathbb{B}^{-1}\begin{pmatrix}\tau^{-\frac{2}{3}}(r_1) \\ \tau^{-\frac{2}{3}}(r)\end{pmatrix}\right] \tag{5.4.2a}$$

where the covariance matrix is

$$\mathbb{B} = \begin{pmatrix} \left\langle \left(\tau^{-\frac{2}{3}}(r_1)\right)^2 \right\rangle & \left\langle \tau^{-\frac{2}{3}}(r_1)\tau^{-\frac{2}{3}}(r) \right\rangle \\ \left\langle \tau^{-\frac{2}{3}}(r_1)\tau^{-\frac{2}{3}}(r) \right\rangle & \left\langle \left(\tau^{-\frac{2}{3}}(r)\right)^2 \right\rangle \end{pmatrix}. \tag{5.4.2b}$$

In writing (5.4.2), we have already used the fact that the mean of $\tau^{-\frac{2}{3}}$ is 0 to simplify the expression somewhat. We will eventually need to obtain the inverse of the covariance matrix $\mathbb{B}$. This is made easier by writing out the inverse of a general 2-by-2 matrix now which is

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}. \tag{5.4.3}$$

We'll further simplify the problem at hand by restricting ourselves specifically to the case of a universe that currently appears to be exactly critical; in other words require $\tau^{-\frac{2}{3}}(r_1) = 0$ with no uncertainty. Returning to (5.4.2a) and using (5.4.3), this allows the joint distribution to be simplified somewhat to

$$P\left(\tau^{-\frac{2}{3}}(r_1) = 0, \tau^{-\frac{2}{3}}(r)\right) = \frac{1}{2\pi\sqrt{\det \mathbb{B}}} \exp\left[-\frac{\left\langle (\tau^{-\frac{2}{3}}(r_1))^2 \right\rangle}{2 \det \mathbb{B}} \left(\tau^{-\frac{2}{3}}(r)\right)^2 \right]. \tag{5.4.4}$$

This is a Gaussian distribution with a mean of 0 and a variance of $\det \mathbb{B} / \left\langle (\tau^{-\frac{2}{3}}(r_1))^2 \right\rangle$ but the distribution is not normalized (the total probability integrated over the one remaining free parameter is not 1). We can integrate this to obtain the cumulative collapse probability as in (5.2.28). This gives

$$P\left(\text{collapse} \in [\tau^{-\frac{2}{3}}(r), \infty)\right) = \frac{1}{\sqrt{8\pi \left\langle (\tau^{-\frac{2}{3}}(r_1))^2 \right\rangle}} \operatorname{erfc}\left(\sqrt{\frac{\left\langle (\tau^{-\frac{2}{3}}(r_1))^2 \right\rangle}{2 \det \mathbb{B}}}\ \tau^{-\frac{2}{3}}(r) \right) \tag{5.4.5}$$

where the determinant of the covariance matrix is

$$\det \mathbb{B} = \left\langle \left(\tau^{-\frac{2}{3}}(r_1)\right)^2 \right\rangle \left\langle \left(\tau^{-\frac{2}{3}}(r)\right)^2 \right\rangle - \left\langle \tau^{-\frac{2}{3}}(r_1)\tau^{-\frac{2}{3}}(r) \right\rangle^2. \tag{5.4.6}$$

Let's also write down, as in Section 5.2.2, the median reduced inverse collapse time which is

$$\operatorname{median}\left(\tau^{-\frac{2}{3}}\right) = \zeta \sqrt{\frac{2 \det \mathbb{B}}{\left\langle (\tau^{-\frac{2}{3}}(r_1))^2 \right\rangle}}. \tag{5.4.7}$$

This is the principle result of this document: the median reduced inverse collapse time of the cosmological material contained in a radius of raH, given that the material in a radius of $r_1 aH$ is known to be critical. Now we just need to compute it.

We already have an expression for the mean square of $\tau^{-\frac{2}{3}}$ in (5.2.26). Having the mean square, all that remains is to find the two-point radial correlation function, $\left\langle \tau^{-\frac{2}{3}}(r_1)\tau^{-\frac{2}{3}}(r) \right\rangle$. We'll follow the same procedure as in Section 5.2.1. Starting from the definition of the reduced inverse collapse time in (5.2.25), the two-point radial correlation function is

$$\left\langle \tau^{-\frac{2}{3}}(r_1)\tau^{-\frac{2}{3}}(r) \right\rangle = \frac{1}{\sigma^2 r_1 r (aH)^2} \left\langle \frac{\Phi'(r_1)}{aH} \frac{\Phi'(r)}{aH} \right\rangle.$$

Using the decomposition of the radial derivative of $\frac{1}{aH}\Phi$ in (5.2.8), we have

$$\begin{aligned} &= \frac{1}{(2\pi)^3 \sigma^2 r_1 r (aH)^4} \left\langle \int k_1 k_2 \, \mathrm{j}_1(k_1 r_1) \, \mathrm{j}_1(k_2 r) \Phi_{k_1} \Phi_{k_2} \, \mathrm{d}^3 k_1 \, \mathrm{d}^3 k_2 \right\rangle \\ &= \frac{1}{(2\pi)^3 \sigma^2 r_1 r (aH)^4} \int k_1 k_2 \, \mathrm{j}_1(k_1 r_1) \, \mathrm{j}_1(k_2 r) \left\langle \Phi_{k_1} \Phi_{k_2} \right\rangle \mathrm{d}^3 k_1 \, \mathrm{d}^3 k_2, \end{aligned}$$

and using (3.2.7) for the two-point spectral correlation function gives

$$= \frac{1}{4\pi\sigma^2 r_1 r (aH)^4} \int \mathrm{j}_1(kr_1) \, \mathrm{j}_1(kr) \mathcal{P}_\Phi(k) \frac{\mathrm{d}^3 k}{k}.$$

Introducing the same smoothing as in (5.2.10) and then doing the angular integration makes this

$$\begin{aligned} &= \frac{2\pi^2}{\sigma^2 r_1 r (aH)^4} \int \mathrm{j}_1(kr_1) \, \mathrm{j}_1(kr) W^2_{k;r_0} \mathcal{P}_\Phi(k) \frac{\mathrm{d}^3 k}{k} \\ &= \frac{(2\pi)^3}{\sigma^2 r_1 r (aH)^4} \int k \, \mathrm{j}_1(kr_1) \, \mathrm{j}_1(kr) W^2_{k;r_0} \mathcal{P}_\Phi(k) \, \mathrm{d}k, \end{aligned} \tag{5.4.8}$$

which, after assuming the same power spectrum as in (5.2.11), finally becomes

$$\left\langle \tau^{-\frac{2}{3}}(r_1)\tau^{-\frac{2}{3}}(r) \right\rangle = \frac{1}{(r_1 aH)} \left(\frac{1}{raH} \right)^{n_{s\Phi}+2} \int x^{n_{s\Phi}} \, \mathrm{j}_1(\frac{r_1}{r} x) \, \mathrm{j}_1(x) \exp\left[-\frac{r_0^2}{r^2} x^2 \right] \mathrm{d}x. \tag{5.4.9}$$

It can be shown that this evaluates to

$$\begin{aligned} \left\langle \tau^{-\frac{2}{3}}(r_1)\tau^{-\frac{2}{3}}(r) \right\rangle = \frac{\sqrt{\pi}}{24} \left(\frac{1}{r_0 aH} \right)^{n_{s\Phi}+3} \times \\ \sum_{m=0}^{\infty} \frac{\Gamma(m + \frac{1}{2}(n_{s\Phi}+3))}{m! \, \Gamma(m+\frac{5}{2})} \left(-\frac{r_1^2}{4r_0^2} \right)^m {}_2\mathrm{F}_1 \left(-m, -\frac{3}{2} - m; \frac{5}{2}; \frac{r^2}{r_1^2} \right) \end{aligned} \tag{5.4.10}$$

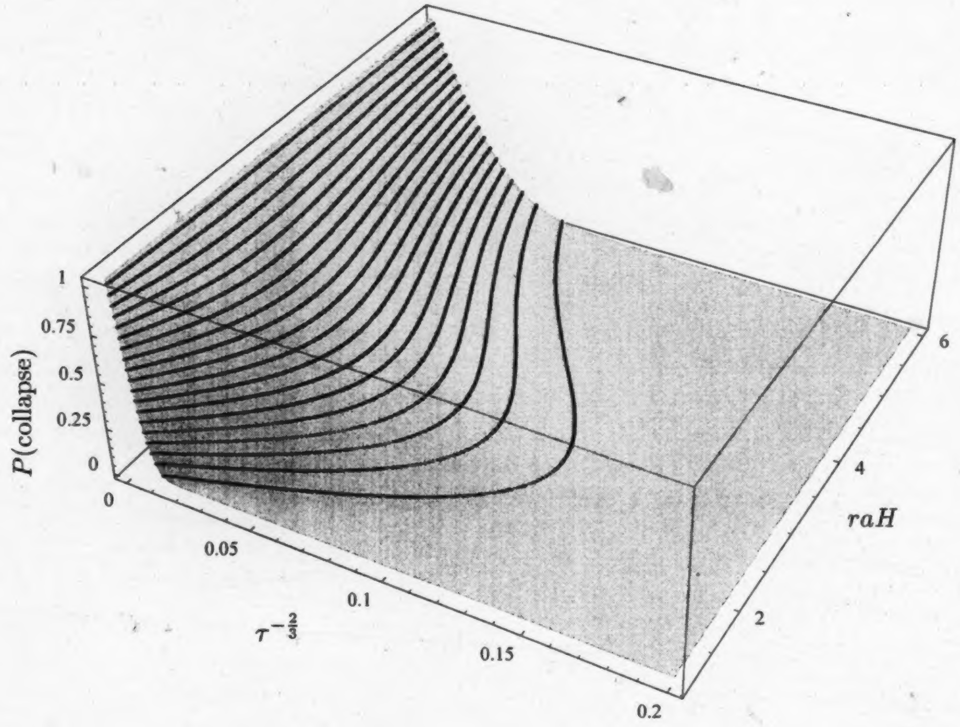


Figure 5.12: The cumulative probability of collapse $P\left(\text{collapse} \in [\tau^{-\frac{2}{3}}(r), \infty)\right)$, conditional on the material inside $r_1 aH$ being critical, as it depends on radius. The contours show from 5% to 95% of the total inclusively in 5% intervals.

for $n_{s\Phi} > -3$. See Appendix C.13 for the derivation.

To visualize the behaviour of the collapse-time distribution, a plot of the cumulative collapse probability in (5.4.5) as a function of the reduced inverse collapse time for a range of radii is shown in Figure 5.12. For definiteness the smoothing radius was set to the Hubble radius, so $r_0 aH = 1$. The observer's current viewable radius has also been set to the Hubble radius, so $r_1 aH = 1$. The plot is specifically for a spectral index of $n_{s\Phi} = 1.01$.

A plot of the median reduced inverse collapse time in (5.4.7) as a function of radius for the present best estimate of the spectral index is shown in Figure 5.13. The same result plotted as a function of elapsed proper time for a centrally-located observer is shown in Figure 5.14. Finally, Figure 5.15 shows in a more exaggerated manner the way in which the behaviour of the median reduced inverse collapse time depends on spectral index. These plots were all made using the same choices of parameters ($r_0 aH$, and $r_1 aH$) as in Figure 5.12.

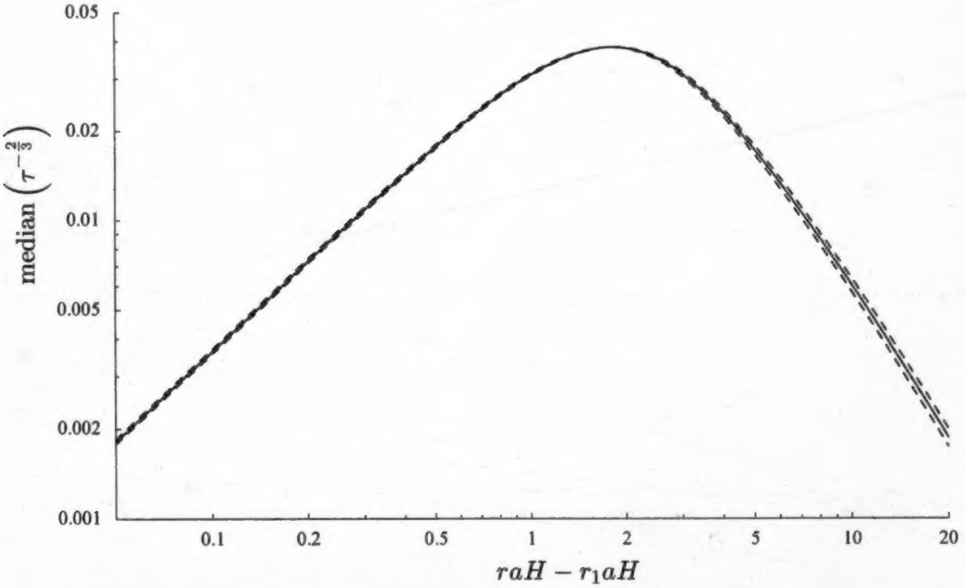


Figure 5.13: A plot of the median reduced inverse collapse time as a function of horizon radius using the current best-estimate for the spectral index. The solid curve is for a spectral index of $n_{s\Phi} = 1.01^{+0.09}_{-0.08}$ with the dashed curves showing the envelope corresponding to the uncertainty range.

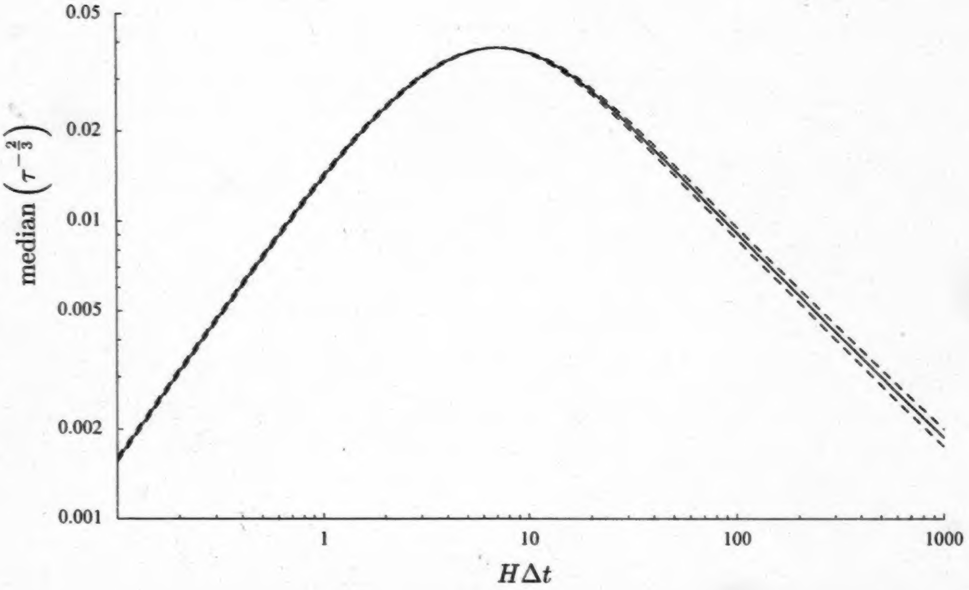


Figure 5.14: A plot of the median reduced inverse collapse time as a function of elapsed proper time for a centrally-located observer using the current best estimate for the spectral index. The solid curve is for a spectral index of $n_{s\Phi} = 1.01^{+0.09}_{-0.08}$ with the dashed curves showing the uncertainty envelope.

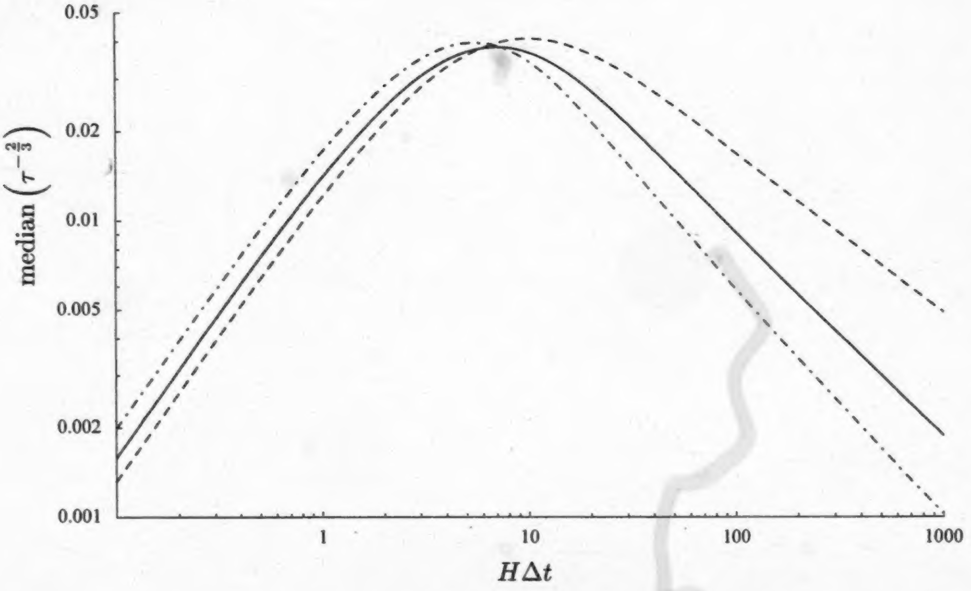


Figure 5.15: A plot of the median reduced inverse collapse time as a function of elapsed proper time for a centrally-located observer for several spectral indices. The solid curve is the same as the solid curve in Figure 5.14 and so shows the result for a spectral index of $n_{s\Phi} = 1.01$. The other two curves show the behaviour of the result for a spectral index of $n_{s\Phi} = 0$ (dashes), and a spectral index of $n_{s\Phi} = 2$ (dots and dashes).

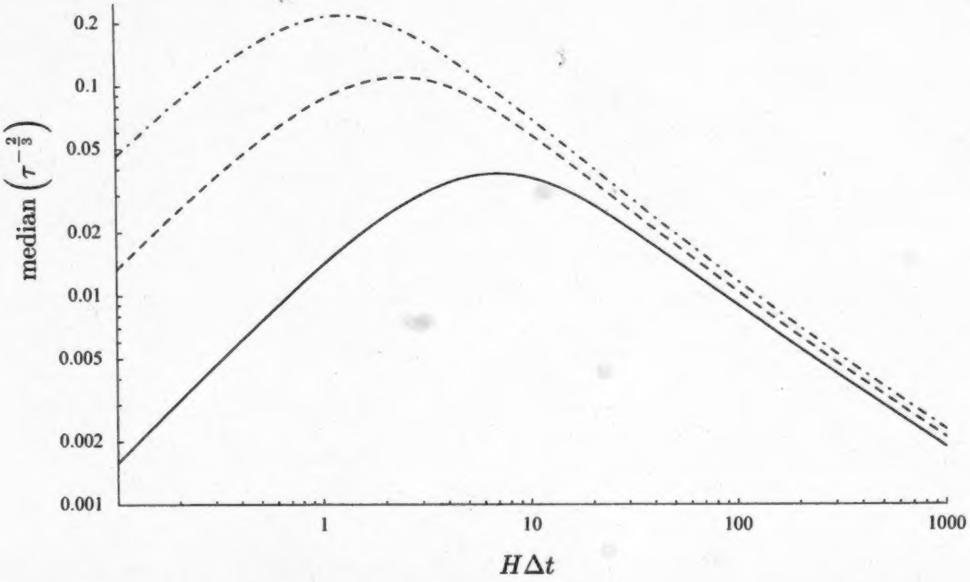


Figure 5.16: The effect of reducing the smoothing radius on the behaviour of the median reduced inverse collapse time. The solid curve is the same solid curve shown in Figure 5.14. The other two curves show the behaviour when the smoothing radius is set to $r_0aH = 1/2$ (dashes) and $r_0aH = 1/4$ (dots and dashes).

The effect of changing the smoothing radius is shown in Figure 5.16. The central curve is made with the same choices of parameters as the solid curve in Figure 5.14. For clarity, the uncertainty envelope from the spectral index has been omitted. Figure 5.17 shows the effect changing the time at which the observer makes their initial measurement has on the results. In particular, it shows three choices of r_1aH, all other parameters being held constant.

By using the cosmological parameters in Table B.5, the spectral amplitude in (5.2.12), and chosing not unreasonable values for the two radial parameters (i.e. 1), and using the relationship between proper collapse time and $\tau^{-\frac{2}{3}}$ in (5.2.24), the axes in Figure 5.14 can be converted to more human-readable units. The result is shown in Figure 5.18 where both the horizontal and vertical axes are in Planck times.

5.4.2 Commentary

The plot in Figure 5.14 is a graphical representation of the principle result of this document: the time evolution of the apparent collapse-time of the visible universe given a single observation of some volume of it being exactly critical. The behaviour of this quantity is governed by two competing processes as time elapses. The first process is the increasing uncertainty about what it is the observer will find as he or she is able to

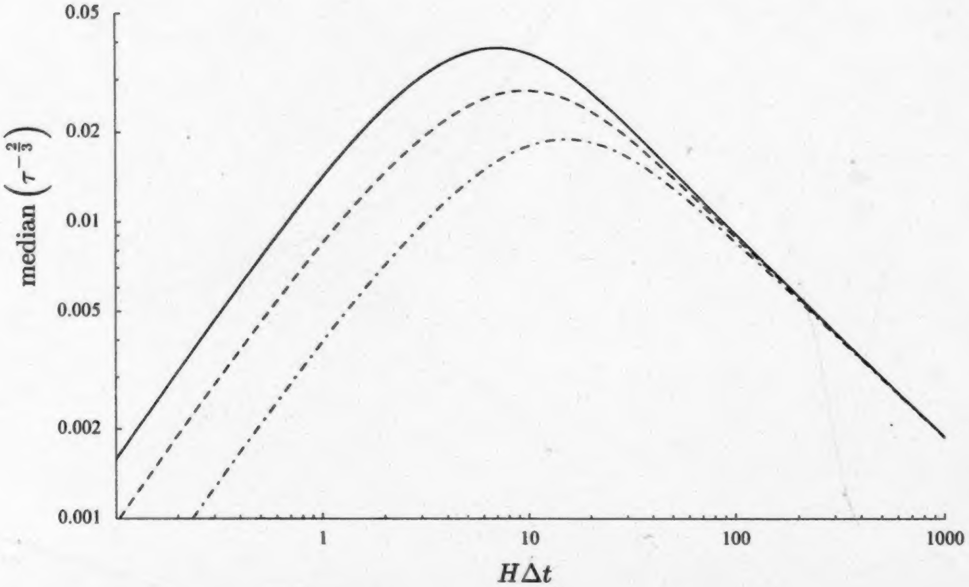


Figure 5.17: The effect of changing the radius of initial observation on the behaviour of the median reduced inverse collapse time. The solid curve is the same solid curve shown in Figure 5.14 for which $r_1aH = 1$. The other two curves show the behaviour when $r_1aH = 2$ (dashes) and $r_1aH = 3$ (dots and dashes).

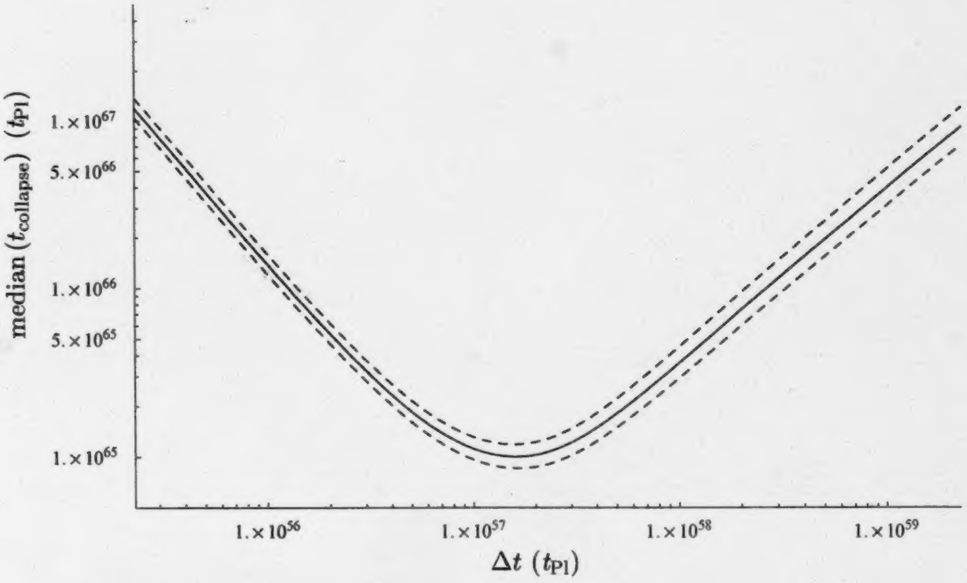


Figure 5.18: A repeat of Figure 5.14 with both axes in Planck times. For this plot, $\delta_{\text{H}} = 1.94 \times 10^{-5} \pm 7\%$ ($\sigma = 2.91 \times 10^{-5} \pm 7\%$), $n_{s\Phi} = 1.01^{+0.09}_{-0.08}$, $H = H_{\text{dec}} = 4.37 \times 10^{-57} \pm 3\%$, $r_0 aH = 1$, and $r_1 aH = 1$. Note: $10^{60}\ t_{\text{Pl}} = 1.7$ Ga.

see out to larger distances. This process increases the median inverse collapse time with time and the increase appears to be as a power of Δt. The second process, which acts to decrease the median with time, is simply the fact that larger things take longer to collapse. This process also appears to be as a power of Δt. Initially, the first process dominates but eventually the second does and the distribution begins to shrink toward a δ-function at $t_{\text{collapse}} = \infty$.

The computation of the median reduced inverse collapse time and its behaviour with respect to the elapsed proper time for a centrally-located observer depend on the values of three parameters: the spectral index of primordial curvature perturbations, $n_{s\Phi}$, the smoothing radius, r_0aH, and the radius of initial observation, r_1aH. In Figure 5.15 we see the effect of changing the spectral index, in Figure 5.16 we see the effect of changing the smoothing radius, and in Figure 5.17 we see the effect of changing the radius of initial observation.

In Figure 5.16 we find that reducing the smoothing radius causes the overall rate of change of the median inverse collapse time to increase. This is seen as a linear upward shift of the curve. We can understand this behaviour by recognizing that by reducing the smoothing radius we are retaining more small-scale structure in the field. With regard to the early-Δt behaviour, increasing the small-scale structure increases the uncertainty about what it is the centrally-located observer will find as their horizon recedes the same fixed distance. Likewise, for the late-Δt behaviour more small-scale fluctuations increase the likelihood of a seeing any given level of mean over-density within the horizon which increases the likelihood of a small collapse time — corresponding to an increase of the median inverse collapse time. This behaviour is consistent with what was seen earlier in Figure 5.7 where we found that the role of the smoothing radius was mostly to adjust the sensitivity of the results to the fluctuations of the underlying field without, otherwise, significantly affecting their behaviour.

A somewhat different result is seen in Figure 5.17 for the effect of the radius of initial observation. In this plot we see that increasing the radius of initial observation has the effect of decreasing the initial rate of increase of the median reduced inverse collapse time — seen as a linear downward shift of the graph to the left of the peak. The volume of initial observation does not, however, appear to have any effect on the late-Δt behaviour. Again, these effects are understandable. By increasing the volume of the universe we initially observe to be exactly critical, we are effectively increasing our knowledge of the general properties of our local part of it so there is less uncertainty about what it is we will see as more of it becomes visible to us. The late-Δt behaviour, however, should be determined by the statistics of large length scales in the underlying field and the physics of the collapse process. Since neither of these can be influenced by our observation, the late-Δt behaviour is insensitive to the choice of r_1aH.

It might be interesting to attempt to remove the late-Δt behaviour from the data: to normalize the collapse time to the typical collapse time at a given radius. Unfortunately, there is no obvious way to do this. The question is that of how to define a “typical” or “characteristic” collapse time. The expected collapse time for all structures is infinite: the background is exactly critical so on average nothing collapses, ever. We’ve even seen that of those regions that are known to eventually collapse the expected collapse time is still infinite. One could, perhaps, compare to the light-crossing time but this is *not* the collapse time so some residual effect will still be seen. One could compare to a power of the light-crossing time, tuned to remove large-$H\Delta t$ asymptotic behaviour but then one comes to the following conclusion: the only good definition of a "typical" collapse time is exactly the result shown in Figure 5.14, and normalizing the outcome to itself is silly.

The spectral index, by far, has the greatest effect on the behaviour of the median inverse collapse time. As seen in Figure 5.15, it changes the exponent in the power-law decay at late-Δt — the only change to either the early- or late-Δt power-law exponents seen when changing any of the parameters. Specifically, reducing the spectral index — increasing the relative fluctuation amplitude at large length scales — has the expected effect of increasing the uncertainty in the collapse time for late times and decreasing it for early times.

It is remarkable that the exponent in what appears to be an early-Δt power-law relationship between the median reduced inverse collapse time and the elapsed proper time for a centrally-located observer is apparently insensitive to all three of the parameters of the computation — the spectral index, the smoothing radius, and the initial volume of observation. Since none of these parameters appear in the relationship between $\tau^{-\frac{2}{3}}$ and t_{collapse} in (5.2.24), it is then the case that the exponent in the early-Δt behaviour of median (t_{collapse}) is itself independent of *all* parameters including the spectral amplitude and the Hubble constant. Numerical fits to the curves suggest that

$$\lim_{H\Delta t\to 0} \frac{\mathrm{d}}{\mathrm{d}\ln(H\Delta t)} \ln \operatorname{median}\left(\tau^{-\frac{2}{3}}\right) \approx 1 \tag{5.4.11}$$

which leads to

$$\lim_{\Delta t\to 0} \frac{\mathrm{d}}{\mathrm{d}\ln(\Delta t)} \ln \operatorname{median}\left(t_{\text{collapse}}\right) \approx -\frac{3}{2}. \tag{5.4.12}$$

Of course, the coefficient in the power-law is not independent of the parameters so the actual numerical value of the median collapse time for any elapsed observation time does depend on all five parameters. Using current cosmological data from Table B.5, and choosing not unreasonable values for the two radial parameters (i.e. 1), Figure 5.18 shows us the behaviour of the median of the distribution for the proper collapse time of the visible universe as a function of the centrally-located observer's elapsed proper time. In this case, the co-efficient for t_{collapse} is about $1.3 \times 10^{150}\ {t_{\text{Pl}}}^{5/2}$ so

$$\operatorname{median}\left(t_{\text{collapse}}\right) \approx \left(1.3 \times 10^{150}\ {t_{\text{Pl}}}^{5/2}\right)(\Delta t)^{-3/2}, \tag{5.4.13}$$

for $\Delta t \leq 10^{57}\ t_{\text{Pl}} \approx 1.7$ Ma. This fit is shown in Figure 5.19. From this:

- If the universe, at a later time, proves to be over-dense and thus on its way to collapse, then approximately 1.7 million years from now we have a 50% chance of observing the material visible to us by that time to have a time-to-collapse of as little as 70 trillion years — depending on the influence of (neglected) small scale structure.

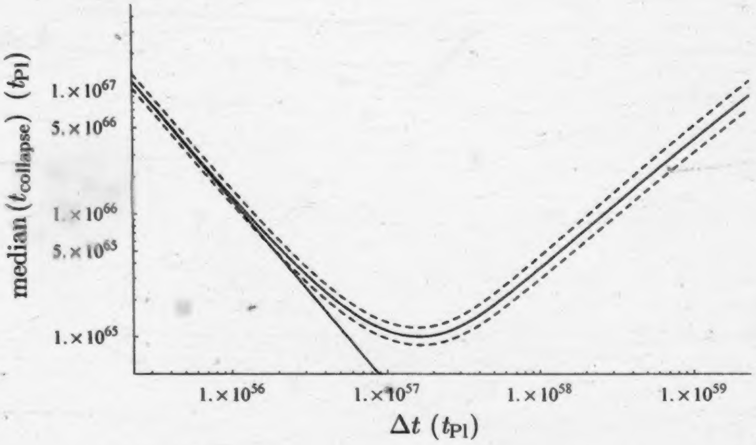


Figure 5.19: A repeat of Figure 5.18 with the polynomial fit median $(t_{\text{collapse}}) = \left(1.3 \times 10^{150}\ {t_{\text{Pl}}}^{5/2}\right)(\Delta t)^{-3/2}$.

5.5 Comparison to the Press-Schechter Model

This section will briefly make the connection between the collapse formalism used in this document and the so-called Press-Schechter model [78].

5.5.1 Overview of the Press-Schechter Model

The result of the Press-Schechter analysis is a distribution of the masses of collapsed objects as a function of red-shift. This information is arrived at by considering an isolated top-hat fluctuation of material in a (in this case) critical-density background. The top-hat fluctuation has a given initial co-moving radius r_i and initial density contrast δ_i. Remarkably, it is found that the red-shift at which such a fluctuation collapses into a compact virialized system depends *only* on the initial density contrast and the red-shift z_i at which it had that contrast [72, equation (8.41)]:

$$(1 + z_{\text{collapse}}) = \frac{1}{(3\pi/2)^{2/3}}\delta_i(1 + z_i). \tag{5.5.1}$$

The final mass of the collapsed object comes from the initial radius of the fluctuation that produces it, the mean density of the universe at the initial time, and the initial density contrast of the fluctuation. All fluctuations with the same initial density contrast collapse at the same time, and their masses vary due to their varying initial sizes. The observed number density of masses at any red-shift then arises from the number density of region sizes in the initial Gaussian fluctuations with the given mean density contrast.

Press-Schechter analysis supposes that at the initial time, the peculiar velocity field for the material in the top-hat fluctuation is 0. This assumption distinguishes standard

Press-Schechter analysis from the collapse statistics studied in this document wherein the initial peculiar velocity field is tuned so as to make the Newtonian potential time-independent. The manner in which this choice enters into the final results is quite straight-forward, however, and a direct comparison can still be made.

5.5.2 Connection to the Formalism of this Document

The first step in connecting the formalism of this document to that of Press-Schechter is to determine the relationship between spherical top-hat fluctuations and the Newtonian potential Φ. We start by defining an over-dense, uniform spherical distribution of dust of initial radius r_i. The initial density contrast of the material shall be described by

$$\delta(\vec{x}) = \begin{cases} \delta_i, & |\vec{x}| \le r_i \\ 0, & |\vec{x}| > r_i, \end{cases} \tag{5.5.2}$$

with $\delta_i > 0$. The Fourier transform of this density contrast is

$$\delta_k = \delta_i r_i^3 \sqrt{\frac{2}{\pi}} \frac{1}{kr_i} \mathrm{j}_1(kr_i). \tag{5.5.3}$$

From (C.9.47) in Appendix C.9.2, the Fourier transform of the Newtonian potential corresponding to this density contrast is

$$\Phi_k = -\delta_i r_i^3 \frac{1}{\sqrt{2\pi}} \frac{1}{\frac{1}{3}\frac{(kr_i)^2}{(r_i aH)^2} + 1} \frac{1}{kr_i} \mathrm{j}_1(kr_i). \tag{5.5.4}$$

In this step, the assumption that $\dot{\Phi} = 0$ has been used. The assumption, instead, that $\delta u_r = 0$ would have involved using (C.9.46) with $U_k = 0$ rather than (C.9.47) to relate the Fourier transform of the density contrast to that of the metric perturbation. This choice would result in the +1 in the denominator of (5.5.4) being absent. Comparison of results to those of Press-Schechter can be made by taking the small-radius limit ($r_i aH \ll 1$) which has the effect of making the presence or otherwise of the +1 insignificant. Continuing by transforming the Newtonian potential back to co-ordinate space, we find

$$\Phi(\vec{x}) = -\delta_i \frac{1}{\pi} \int \frac{kr_i}{\frac{1}{3}\frac{(kr_i)^2}{(r_i aH)^2} + 1} \mathrm{j}_1(kr_i)\, \mathrm{j}_i(kr)\, \mathrm{d}(kr_i). \tag{5.5.5}$$

Using the recurrence relation in (5.2.6), we have $\frac{1}{aH}\frac{\mathrm{d}}{\mathrm{d}r}\mathrm{j}_i(kr) = -\frac{kr}{raH}\mathrm{j}_1(kr)$ so the radial derivative of the Newtonian potential is

$$\frac{\Phi'}{aH} = \frac{\delta_i}{raH} \frac{1}{\pi} \int \frac{(kr_i)(kr)}{\frac{1}{3}\frac{(kr_i)^2}{(r_i aH)^2} + 1} \mathrm{j}_1(kr_i)\, \mathrm{j}_1(kr)\, \mathrm{d}(kr_i), \tag{5.5.6}$$

and evaluating this at the outer edge of the top-hat, $r = r_i$, yields

$$\frac{1}{aH}\Phi'(r_i aH) = \frac{\delta_i}{r_i aH}\frac{1}{\pi}\int_0^\infty \frac{x^2}{\frac{1}{3}\frac{x^2}{(r_i aH)^2} + 1}\,\mathrm{j}_1^2(x)\,\mathrm{d}x \tag{5.5.7}$$

where x is a dummy integration variable. This result is the bridge between the Press-Schechter formalism, in which one considers the statistics of top-hat fluctuations of given initial radii and density contrasts, and the formalism of this document in which one considers the statistics of the radial derivative of the Newtonian potential. Through (5.5.7), one can map statistical information from one formalism to the other: at a radius r_i, this expression gives the mapping between a top-hat fluctuation of amplitude δ_i and a Newtonian potential with radial derivative Φ'.

5.5.3 Collapse-Time for a Spherical Top-Hat Fluctuation

To demonstrate, explicitly, the equivalence of the formalism of this document to that of Press-Schechter, (5.5.1) will be re-derived from the expressions used in this document. In so doing, the demonstration of the agreement of the collapse model of this document with observation is "piggy-backed" onto the existing body of research demonstrating the agreement between the Press-Schechter model and observation.

Using the radial derivative of Φ' from (5.5.7) in (5.1.78) gives the collapse time for a top-hat fluctuation with initial density contrast δ_i and initial radius $r_i aH$,

$$t_{\text{collapse}} = \pi H^{-1}\sqrt{\frac{3\pi}{10\delta_i}}^{\,3}(r_i aH)^3\left(\int_i^\infty \frac{x^2}{\frac{1}{3}\frac{x^2}{(r_i aH)^2} + 1}\,\mathrm{j}_1^2(x)\,\mathrm{d}x\right)^{-3/2}. \tag{5.5.8}$$

Here, again, the assumption that $\dot{\Phi} = 0$ has been used, since it is implicit in (5.1.78). As discussed in Section 5.1.5, the Press-Schechter assumption, that instead $\delta u_r = 0$, has the effect of reducing the collapse time by a factor of $(3/5)^{3/2}$. The behaviour of the collapse time in (5.5.8) is shown in Figure 5.20, normalized to the initial density contrast, δ_i. The small-r_i limit for the collapse time can be found by writing the integral in (5.5.8) as

$$3(r_i aH)^2\int_i^\infty \frac{x^2}{x^2 + 3(r_i aH)^2}\,\mathrm{j}_1^2(x)\,\mathrm{d}x \approx 3(r_i aH)^2\int_i^\infty \mathrm{j}_1^2(x)\,\mathrm{d}x = (r_i aH)^2\frac{\pi}{2},$$

which makes the collapse time

$$t_{\text{collapse}}(r_i aH \ll 1) = \pi H^{-1}\sqrt{\frac{3}{5\delta_i}}^{\,3} = t_i\pi\frac{3}{2}\sqrt{\frac{3}{5\delta_i}}^{\,3}, \tag{5.5.9}$$

where (2.1.13) has been used. Here, we have recovered the essential feature of the Press-Schechter formalism: that the time of collapse depends only the fluctuation's initial density contrast. For a flat universe, the collapse time can be converted into a red-shift

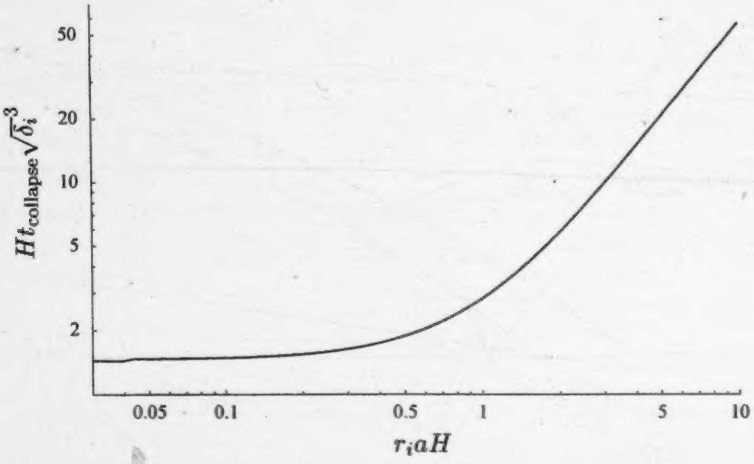


Figure 5.20: The collapse time Ht_{collapse} (in Hubble times) for a spherical top-hat fluctuation of initial proper radius $r_i aH$ (in Hubble radii) in an expanding, critical-density, background normalized to the fluctuation's initial density contrast, δ_i. Note the asymptotically flat behaviour at small radii.

using the relationship [72, equation (2.82)]

$$\frac{1+z_{\text{collapse}}}{1+z_i} = \left(\frac{t_i}{t_i + t_{\text{collapse}}}\right)^{\frac{2}{3}}. \tag{5.5.10}$$

We assume that $\delta_i \ll 1$, and therefore that $t_{\text{collapse}}/t_i \gg 1$, which yields

$$\frac{1+z_{\text{collapse}}}{1+z_i} \approx (t_{\text{collapse}}/t_i)^{-\frac{2}{3}} = \frac{1}{(3\pi/2)^{2/3}}\delta_i\frac{5}{3},$$

or

$$(1+z_{\text{collapse}}) = \frac{5}{3}\frac{1}{(3\pi/2)^{2/3}}\delta_i(1+z_i). \tag{5.5.11}$$

Compare this result to (5.5.1). The additional factor of 5/3 in this result is exactly the additional factor present in the expression for the collapse time when the initial peculiar velocity field is chosen so as to make $\dot{\Phi} = 0$ rather than $\delta u_r = 0$. Otherwise, this demonstrates the equivalence of the collapse model used in this document to that of the Press-Schechter formalism.

Chapter 6

Conclusions

6.1 Summary

The computation we have now completed has shown us how it is that the apparent collapse time for the universe may evolve over time given our current knowledge of the universe's parameters. Let me summarize the steps taken to obtain this.

1. Introduce a "background" description for the cosmos: the $k = 0$ Friedmann-Robertson-Walker cosmological model.
2. Place, alternately, a homogeneous scalar field (the inflaton) and then dust in the model to obtain the time-evolution of the background.
3. Introduce linear-order perturbations to the inflaton field, the cosmological metric, and the dust to obtain a model for perturbations to the cosmological model. The conformal Newtonian gauge was chosen for this description.
4. Use the perturbation model to justify extending the presently-observed fluctuation spectrum to super-horizon scales.
5. Construct a model for the radial collapse of a spherically symmetric cosmological perturbation. Use this model to obtain an expression for the collapse time of a fluctuation from the model's initial conditions.
6. Express the spherical collapse model's initial conditions in terms of the parameters of the linear perturbations.
7. Use the statistics for the parameters of the linear perturbations (power spectrum, etc.) to obtain distributions for the spherical collapse model's initial conditions and thus distributions for the collapse time of shells of different radii.
8. Re-express the spherical collapse model's radial co-ordinate in terms of the length of time required for a centrally-located observer to see out that far based on the increasing size of a hypothetical horizon, and thus obtain the time evolution of the distribution of collapse times as perceived by the centrally-located observer.

The most concise expression of the result of this analysis is the value of the median proper collapse time of the visible universe, $\mathrm{median}\,(t_{\mathrm{collapse}})$, as a function of elapsed proper time for a centrally-located observer, Δt. The computation of this function requires the specification of five parameters:

- the spectral index of the primordial curvature perturbations, $n_{s\Phi}$,
- the amplitude of the primordial curvature perturbation spectrum, σ^2,
- the Hubble constant on the observation-limiting horizon, H,
- the smoothing radius chosen to remove small-scale structure from the primordial perturbations, $r_0 aH$, and
- the radius out to which the observer is initially able to see, $r_1 aH$, (equivalent to the time since the observation-limiting horizon at which the initial observation is made).

A median "reduced inverse collapse time," $\mathrm{median}\left(\tau^{-\frac{2}{3}}\right)$, can be defined that is independent of the amplitude of the curvature spectrum and the Hubble constant on the horizon, and is linear in the primordial curvature perturbations. The algebraic dependence of this quantity on the three remaining parameters is quite complex, however, so no further isolation of the parameters has been achieved at this time. A numerical exploration of the behaviour of this function suggests a power-law behaviour for early and late times. The exponent for the early-time behaviour appears to be independent of all the parameters of the analysis while the exponent for the late-time behaviour is sensitive to the spectral index.

6.2 Recent Revisions of Cosmological Parameters

Since the work in this document was begun, the best estimates for various cosmological parameters have changed several times. The values used throughout this document are quoted in Appendix B but recently, with the publication of the results from the first year of observations of the Wilkinson Microwave Anisotropy Probe (WMAP), the best fit values for these parameters have been revised again. The results of these new observations can be found in [10], [49], [73], [76], and [84]. Since the latest values are not significantly different from the ones used in this document, I have decided to not re-compute the results of this document using the new values. For comparison, the new values from the WMAP are shown in Table 6.1. These are the across-the-board best fit values to the data. In [76] are presented best fit values for various classes of inflation model. There, one finds that the WMAP data rule out the $\frac{1}{4}\lambda\phi^4$ inflaton potential at more than three sigma.[1] The data do support a single-field $\frac{1}{2}m^2\phi^2$ inflation model but when this model is assumed, the best-fit values for the cosmological parameters are slightly different than those of Table 6.1.

---

[1] See [76, Section 3.4.1].

Hubble parameter [10, Table 3]	$h = 0.71^{+0.04}_{-0.03}$
Matter density [10, Table 3]	$\Omega_{\mathrm{m}}h^2 = 0.135^{+0.008}_{-0.009}$
Baryon density [10, Table 3]	$\Omega_{\mathrm{b}}h^2 = 0.0224 \pm .0009$
Dark energy density [10, Table 3]	$\Omega_\Lambda = 0.73 \pm 0.04$
Total density [10, Table 3]	$\Omega_0 = 1.02 \pm .02$
Spectral index [10, Table 3]	$n_{\mathrm{s}}(k_0 = 0.05\ \mathrm{Mpc}^{-1}) = 0.93 \pm 0.03$
Spectral index slope [10, Table 3]	$\mathrm{d}n_{\mathrm{s}}/\mathrm{d}\ln k = -0.031^{+0.016}_{-0.018}$

Table 6.1: Revised cosmological parameters from the WMAP.

The implications for this document are (i) that all the stuff about $\frac{1}{4}\lambda\phi^4$ potentials could be deleted and (ii) that the extrapolation of the fluctuation spectrum to large radii is solidly justified, since it was with the $\frac{1}{2}m^2\phi^2$ model that this was a particularly reasonable thing to do. Unfortunately, the observations do strongly suggest that a pure power-law form for the spectrum is inappropriate: a slope is observed for n_{s} that is inconsistent with 0 at about the two-sigma level. Incorporating this into the computations performed in this document is very difficult. None of the integrals involved in the statistical computations in Chapter 5 can be done analytically with this type of power spectrum so potentially difficult numerical computations would need to be performed in order to obtain revised results.

6.3 Possible Extensions to this Work

It is always difficult, in the end, to bring an investigation to a point where it is undeniably "finished." The work presented here is certainly not an exception so before closing, it is worth commenting on some of the ways an interested investigator might continue, improve, or expand upon it. There have been many places in this document where approximations have been made in order to simplify the computations. These make obvious locations where more accurate approaches could be used to improve or generalize the results. These include the following.

Generalize the metric perturbation's spectrum away from a pure power-law form

As mentioned in Section 6.2, the best current observational data suggests that the spectral index is not independent of k. All of the computations in this document, however, assume that the spectral index is a constant. Weakening this assumption is a straightforward generalization but a tedious one. It would be necessary to recompute the integrals in (5.2.13), (5.3.8), and (5.4.8) and propagate the modifications to all results that depend on them. Unfortunately, with a spectral index that depends on k, the integrals cannot be done analytically so some sort of numerical approximation technique would need to be devised.

Generalize the equation of state

In several places throughout the collapse-time computation, the equation of state for the cosmological fluid has been used. This enters via the equations of motion for the cosmological perturbations. For the bulk of this document, the equation of state has been assumed to be that of dust. This is clearly an invalid assumption based on current observations and thus makes a perfect place for improvements to be made. Ideally one would consider a two-component fluid consisting of dust and a cosmological constant.

Determine the collapse time observed at the centre rather than edge of a fluctuation

The collapse time computed in Section 5.1 is the proper time that elapses along the world line of a shell between any given event on that world line and the shell's eventual collapse. This is *not* the proper time that elapses along the world line of an observer at the centre of the shell. It is the latter that is more relevant to us since we are at the centre of what it is we can observe. There are two things that contribute to the shell's clock and the centre's clock running at a different pace:

- the inward collapse velocity of the shell will cause its clock to appear to be running slow to an observer at the centre while
- the larger the over-density of matter enclosed by the shell, the greater the gravitational blue-shift of its clock as seen at the centre.

Both effects could be accounted for through an appropriate analysis of the Tolman-Lemaître space-time geometry.

Include cosmological perturbations when mapping an elapsed time to an horizon increase

The computation that was done to obtain (5.4.1) was only carried out to zeroth-order in the perturbations. This has the effect of making it appear as though a cosmological fluctuation could land on the head of an observer at its centre before he or see has seen it coming. It is easily seen that the calculations suggest this by noting that the cumulative collapse probability for any given radius is non-zero for all non-zero collapse times — even collapse times less than the time that needs to elapse before a centrally-located observer can see out to that radius.

Allow for shell-crossing

The co-ordinate system is co-moving so if matter from an outer radius falls onto matter at a smaller radius, the co-ordinate system becomes singular. This property of the co-ordinate system required us to stipulate that the collapse time be a monotonically-increasing function of radius (although nothing much was made of this constraint). One possible way to weaken this requirement might be to periodically "reset" the co-ordinate system — to slice space-time up into slabs of some time interval and construct improved co-ordinates at each interface. The effect of this would be to shuffle matter fluctuations around in radius: sometimes mass will be "handed-off" to a larger radial co-ordinate and sometimes to a smaller one in order to prevent the co-ordinate system from becoming singular. The total mass enclosed by any radial co-ordinate thus becomes a Markov process and in the limit of the time interval between co-ordinate resets being taken to zero it becomes a continuous Markov process. The analysis of this process should allow one to obtain the random field that is the enclosed mass as a function of proper radius, as well as the field's time evolution. This information could allow one to answer questions such as "what is the probability of there being a Schwarzschild horizon between r and $r + \mathrm{d}r$ as a function of time?"

This procedure may need to be performed numerically, and the "reseting" mentioned above is equivalent to imposing co-ordinate conditions on the numerical solver. There is a large body of literature on the use of co-ordinate conditions in numerical relativity; for an overview, see [51].

More Exotic Possibilities

In addition to the direct extensions of this work listed above, one could instead consider other collapse mechanisms altogether. The world line of an observer need not end in a space-time singularity caused by the collapse of the matter in their own part of the universe. In [86], Starobinsky takes an entertaining look at some of the more exotic possibilities by which an observer's world line may terminate and it might be possible to assign believable probabilities per unit time to some of them. For example, an observer could collide with a space-time singularity with infinite values of the Riemann tensor concentrated on a null hyper-surface — a gravitational shock wave. These can be propagating through space-time to hit an observer without warning, and are conjectured to be a feature of space-times containing charged or rotating black holes[2] (which ours most certainly does). Starobinsky also explains that the "dark energy" driving the current period of inflation might itself cause the rapid *collapse* of regions of space. If the field's potential contains a term coupling it to gravity of the form $\xi R\phi^2$, and ϕ passes through the critical value of $1/\sqrt{8\pi\xi G}$, the effective gravitational constant becomes infinite causing the smallest of spatial inhomogeneities to rapidly grow without limit. And, at last, it is also possible that one simply collides with a spontaneous singularity caused by random quantum-gravitational fluctuations. Starobinsky's back-of-the-envelope calculation gives a probability for this occurring per unit time of $\sim \mathrm{e}^{-10^{122}}$ — a fantastically small probability but one that is *not* 0 so if nothing else gets you, this must.

---

[2] See [77].

Bibliography

[1] Elcio Abdalla and Roya Mohayaee. Gravitational clustering to all perturbative orders. *Brazilian Journal of Physics*, 31(1):42–44, March 2001, arXiv:astro-ph/9811119.

[2] Milton Abramowitz and Irene A. Stegun, editors. *Handbook of Mathematical Functions.* Dover Publications, Inc., 9th edition, 1970.

[3] George Brown Arfken and Hans Jurgen Weber. *Mathematical Methods for Physicists.* Academic Press, Inc., 4th edition, 1995.

[4] David Bailin and Alexander Love. *Introduction to Gauge Field Theory.* Institute of Physics Publishing, revised edition, 1993.

[5] James M. Bardeen. Gauge-invariant cosmological perturbations. *Physical Review*, D22(8):1882–1905, 1980.

[6] James M. Bardeen. Cosmological perturbations from quantum fluctuations to large scale structure. In Li-Zhi Fang and A. Zee, editors, *Particle Physics and Cosmology: Lectures given at 2nd Guo Shou-jing Summer School on Particle Physics and Cosmology, Nanjing, China, July 1988*, volume 5 of *China Center for Advanced Science and Technology (World Laboratory) Symposium / Workshop Proceedings.* Gordon and Breach, 1989. Also available as DOE/ER/40423-01-C8 and CONF-8807143-1.

[7] James M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay. The statistics of peaks of Gaussian random fields. *The Astrophysical Journal*, 304:15–61, May 1986.

[8] James M. Bardeen, Paul J. Steinhardt, and Michael S. Turner. Spontaneous creation of almost scale-free density perturbations in an inflationary universe. *Physical Review*, D28(4):679–693, August 1983.

[9] John D. Barrow and Frank J. Tipler. *The Anthropic Cosmological Principle.* Oxford University Press, 1986.

[10] C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, D. N. Spergel, G. S. Tucker, E. Wollack, E. L. Wright, C. Barnes, M. R. Grearson, R. S. Hill, E. Komatsu, M. R. Nolta, N. Odegard, H. V. Peiris, L. Verde, and J. L. Weiland. First year Wilkinson Microwave Anisotropy Probe

(WMAP) observations: Preliminary maps and basic results. *The Astrophysical Journal Supplement Series*, 148(1):1–27, September 2003, arXiv:astro-ph/0302207.

[11] Lars Bergström and Ariel Goobar. *Cosmology and Particle Astrophysics.* John Wiley & Sons, Ltd., 1999.

[12] N. D. Birrell and P. C. Davies. *Quantum Fields in Curved Space.* Cambridge University Press, 1982.

[13] Jelle P. Boersma. On the large-scale structure of the universe. August 1998, arXiv:gr-qc/9808074.

[14] S. Bonazzola, E. Gourgoulhon, and J.-A. Marck. Spectral methods in general relativistic astrophysics. *Journal of Computational and Applied Mathematics*, 109(1–2):433–473, September 1999, arXiv:gr-qc/9811089.

[15] I. N. Bronshtein and K. A. Semendyayev. *Handbook of Mathematics.* Van Nostrand Reinhold Company, Inc., 1985. English translation edited by K. A. Hirsch.

[16] T. S. Bunch and P. C. W. Davies. Quantum field theory in de Sitter space: Renormalization by point-splitting. *Proceedings of the Royal Society of London*, A360(1700):117–134, 1978.

[17] Samuel Byland and David Scialom. Evolution of the Bianchi type I, Bianchi type III, and the Kantowski-Sachs universe: Isotropization and inflation. *Physical Review*, D57(10):6065–6074, May 1998, arXiv:gr-qc/9802043.

[18] Esteban Calzetta and B. L. Hu. Correlations, decoherence, dissipation, and noise in quantum field theory. January 1995, arXiv:hep-th/9501040.

[19] Esteban Calzetta and B. L. Hu. Quantum fluctuations, decoherence of the mean field, and structure formation in the early universe. *Physical Review*, D52(12):6770–6788, December 1995, arXiv:gr-qc/9505046.

[20] Peter Coles and Francesco Lucchin. *Cosmology: The Origin and Evolution of Cosmic Structure.* John Wiley & Sons, Ltd., second edition, 2002.

[21] N. Deruelle and D. S. Goldwirth. Conditions for inflation in an initially inhomogeneous universe. *Physical Review*, D51(4):1563–1568, February 1995.

[22] Jaan Einasto, Maret Einasto, Erik Tago, Alexei A. Starobinsky, Fernando Atrio-Barandela, Volker Müller, Alexander Knebe, and Renyue Cen. Steps toward the power spectrum of matter. iii. the primordial spectrum. *Astrophysical Journal*, 519(2):469–478, July 1999, arXiv:astro-ph/9812249.

[23] Zuhui Fan and James M. Bardeen. Distribution of Fourier modes of cosmological density fields. *Physical Review*, D51(12):6714–6721, June 1995, arXiv:astro-ph/9505017.

[24] D. J. Fixsen, E. S. Cheng, J. M. Gales, J. C. Mather, R. A. Shafer, and E. L. Wright. The cosmic microwave background spectrum from the full COBE FIRAS data set. *The Astrophysical Journal*, 473(2):576–587, December 1996.

[25] Crispin W. Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. Springer-Verlag, 1983.

[26] D. S. Goldwirth. Inhomogeneous initial conditions for inflation. *Physical Review*, D43(10):3204–3213, May 1991.

[27] D. S. Goldwirth and T. Piran. Spherical inhomogeneous cosmologies and inflation — numerical-methods. *Physical Review*, D40(10):3263–3279, November 1989.

[28] Izrail Solomonovich Gradshtein and Iosif Moiseevich Ryzhik. *Table of Integrals, Series, and Products*. Academic Press, 5th edition, 1994.

[29] Ø. Grøn. Repulsive gravitation and inflationary universe models. *American Journal of Physics*, 54(1):46–52, January 1986.

[30] Alan H. Guth. Inflationary universe: A possible solution to the horizon and flatness problems. *Physical Review*, D23(2):347–356, January 1981.

[31] Alan H. Guth and So-Young Pi. Fluctuations in the new inflationary universe. *Physical Review Letters*, 49(15):1110–1113, October 1983.

[32] Alan H. Guth and So-Young Pi. Quantum mechanics of the scalar field in the new inflationary universe. *Physical Review*, D32(8):1899–1920, October 1985.

[33] E. R. Harrison. Fluctuations at the threshold of classical cosmology. *Physical Review*, D1(10):2726–2730, May 1970.

[34] Barry R. Holstein. *Topics In Advanced Quantum Mechanics*. Advanced book program. Addison-Wesley, 1992.

[35] Michael A. Hoskin, editor. *The Cambridge Illustrated History of Astronomy*. Cambridge Illustrated History. Cambridge University Press, 1997.

[36] Edwin P. Hubble. Extra-galactic nebulae. *Astrophysical Journal*, 64:321–369, 1926.

[37] Werner Israel. Singular hypersurfaces and thin shells in general relativity. *Il Nuovo Cimento*, 44 B(1), July 1966. See also errata in [38].

[38] Werner Israel. Singular hypersurfaces and thin shells in general relativity. *Il Nuovo Cimento*, 48 B(2), April 1967.

[39] A. V. Ivanov and N. N. Leonenko. *Statistical Analysis of Random Fields*, volume 28 of *Mathematics and Its Applications (Soviet Series)*. Kluwer Academic, english edition, 1989. Translated from the Russian by S. Kotz.

[40] John David Jackson. *Classical Electrodynamics.* John Wiley & Sons, Ltd., 3rd edition, 1999.

[41] D. Kannan. *An Introduction to Stochastic Processes.* Elsevier North Holland, 1979.

[42] Claus Kiefer and David Polarski. Emergence of classicality for primordial fluctuations: Concepts and analogies. *Annalen Der Physik*, 7(3):137–158, 1998, arXiv:gr-qc/9805014.

[43] Claus Kiefer, David Polarski, and Alexei A. Starobinsky. Quantum-to-classical transition for fluctuations in the early universe. *International Journal of Modern Physics*, D7(3):455–462, June 1998.

[44] Lev Kofman, Andrei Linde, and Alexei A. Starobinski. Reheating after inflation. *Physical Review Letters*, 73(24):3195–3198, December 1994, arXiv:hep-ph/9405187.

[45] Lev Kofman, Andrei Linde, and Alexei A. Starobinski. Towards the theory of reheating after inflation. *Physical Review*, D56(6):3258–3295, September 1997, arXiv:hep-ph/9704452.

[46] Lev A. Kofman. The origin of matter in the universe: Reheating after inflation. In Bernard J. T. Jones and Dragoljub Markovic, editors, *Relativistic Astrophysics*. Cambridge University Press, June 1997, arXiv:astro-ph/9605155. Proceedings of a Conference in Honour of Igor Novikov's 60th Birthday.

[47] Lev A. Kofman, V. F. Mukhanov, and Dmitri Yu. Pogosyan. Evolution of inhomogeneities in inflationary models in a theory of gravitation with higher derivatives. *Soviet Physics JETP*, 66(3):433–440, September 1987, arXiv:gr-qc/9805014.

[48] Edward W. Kolb and Michael S. Turner. *The Early Universe.* Addison-Wesley, 1990.

[49] E. Komatsu, A. Kogut, M. R. Nolta, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, M. Limon, S. S. Meyer, L. Page, D. N. Spergel, G. S. Tucker, L. Verde, E. Wollack, and E. L. Wright. First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Tests of Gaussianity. *The Astrophysical Journal Supplement Series*, 148(1):119–134, September 2003, arXiv:astro-ph/0302223.

[50] Lawrence M. Krauss. A new cosmological paradigm: the cosmological constant and dark matter. July 1998, arXiv:hep-ph/9807376.

[51] Luis Lehner. Numerical relativity: a review. *Classical and Quantum Gravity*, 18(17):R25–R86, September 2001, arXiv:gr-qc/0106072.

[52] Abbé Georges Lemaître. L'univers en expansion. *Annales de la Société Scientifique de Bruxelles*, A53:51, 1933.

[53] Abbé Georges Lemaître. The expanding universe. *General Relativity and Gravitation*, 29(5):641–680, May 1997. Translated from the original, [52], by M. A. H. MacCallum.

[54] Julien Lesgourgues, David Polarski, and Alexei A. Starobinsky. Quantum-to-classical transition of cosmological perturbations for non-vacuum initial states. *Nuclear Physics*, B497(1–2):479–508, July 1997, arXiv:gr-qc/9611019.

[55] Andrew R. Liddle and David H. Lyth. *Cosmological Inflation and Large-Scale Structure.* Cambridge University Press, 2000.

[56] Andrew R. Liddle, David H. Lyth, R. K. Schaefer, and Q. Shafi. Pursuing parameters for critical-density dark matter models. *Monthly Notices of the Royal Astronomical Society*, 281(2):531–551, July 1996, arXiv:astro-ph/9511057.

[57] Allan P. Lightman, William H. Press, Richard H. Price, and Saul A. Teukolsky. *Problem Book in Relativity and Gravitation.* Princeton University Press, 1975.

[58] Andrei Linde. A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems. *Physics Letters*, B108(6):389–393, 1982.

[59] Andrei Linde. Chaotic inflation. *Physics Letters*, B129(3–4):177–181, 1983.

[60] Andrei Linde. *Particle Physics and Inflationary Cosmology*, volume 5 of *Contemporary Concepts in Physics.* Harwood Academic Publishers, 1990. Translated from the Russian by Marc Damashek.

[61] Abraham Loeb. Cosmological formation of quasar black holes. *The Astrophysical Journal*, 403(2):542–551, February 1993.

[62] Hugo Martel and Paul R. Shapiro. The asymptotic collapsed fraction in an eternal universe. March 1999, arXiv:astro-ph/9903425. Submitted to Monthly Notices of the Royal Astronomical Society, March 1999.

[63] Hugo Martel, Paul R. Shapiro, and Steven Weinberg. Likely values of the cosmological constant. *The Astrophysical Journal*, 492(1):29–40, January 1998, arXiv:astro-ph/9701099.

[64] Alessandro Melchiorri and Louise M. Griffiths. From anisotropy to omega. November 2000, arXiv:astro-ph/0011147.

[65] Charles W. Misner, Kip S. Thorne, and John A. Wheeler. *Gravitation.* W. H. Freeman and Company, 1973.

[66] Peter J. Mohr and Barry N. Taylor. CODATA recommended values of the fundamental physical constants: 1998. *Journal of Physical and Chemical Reference Data*, 28(6):1713–1852, November–December 1999. Also available in [67].

[67] Peter J. Mohr and Barry N. Taylor. CODATA recommended values of the fundamental physical constants: 1998. *Reviews of Modern Physics*, 72(2):351–495, April 2000.

[68] V. F. Mukhanov. Gravitational instability of the universe filled with a scalar field. *JETP Letters*, 41(9):493–496, 1985.

[69] V. F. Mukhanov, H. A. Feldman, and R. H. Brandenberger. Theory of cosmological perturbations. *Physics Reports*, 215(5–6):203–333, June 1992.

[70] Jayant Vishnu Narlikar. *An Introduction to Cosmology.* Cambridge University Press, third edition, 2002.

[71] Keith A. Olive. Primordial nucleosynthesis and dark matter. July 1997, arXiv:astro-ph/9707212.

[72] Thanu Padmanabhan. *Structure Formation in the Universe.* Cambridge University Press, 1993.

[73] L. Page, M. R. Nolta, C. Barnes, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut M. Limon, S. S. Meyer, H. V. Peiris, D. N. Spergel, G. S. Tucker, E. Wollack, and E. L. Wright. First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Interpretation of the TT and TE angular power spectrum peaks. *The Astrophysical Journal Supplement Series*, 148(1):233–241, September 2003, arXiv:astro-ph/0302220.

[74] Phillip James Edwin Peebles. *The Large-Scale Structure of the Universe.* Princeton University Press, 1980.

[75] Phillip James Edwin Peebles. *Principles of Physical Cosmology.* Princeton University Press, 1993.

[76] H. V. Peiris, E. Komatsu, L. Verde, D. N. Spergel, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, G. S. Tucker, E. Wollack, and E. L. Wright. First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Implications for inflation. *The Astrophysical Journal Supplement Series*, 148(1):213–231, September 2003, arXiv:astro-ph/0302225.

[77] Eric Poisson and Werner Israel. Internal structure of black holes. *Physical Review*, D41(6):1796–1809, March 1990.

[78] Walter H. Press and Paul Schechter. Formation of galaxies and clusters of galaxies by self-similar gravitational condensation. *The Astrophysical Journal*, 187(3):425–438, February 1974.

[79] Adam G. Riess, Alexei V. Filippenko, Peter Challis, Alejandro Clocchiatti, Alan Diercks, Peter M. Garnavich, Ron L. Gilliland, Craig J. Hogan, Saurabh Jha, Robert P. Kirshner, B. Leibundgut, M. M. Phillips, David Reiss, Brian P.

Schmidt, Robert A. Schommer, R. Chris Smith, J. Spyromilio, Christopher Stubbs, Nicholas B. Suntzeff, and John Tonry. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *Astronomical Journal*, 116(3):1009–1038, September 1998, arXiv:astro-ph/9805201.

[80] D. S. Salopek. Generating non-Gaussian adiabatic fluctuations from inflation. March 1999, arXiv:astro-ph/9903327.

[81] Bernard F. Schutz. *A First Course in General Relativity.* Cambridge University Press, 1990.

[82] J. L. Sievers, J. R. Bond, J. K. Cartwright, C. R. Contaldi, B. S. Mason, S. T. Myers, S. Padin, T. J. Pearson, U.-L. Pen, D. Pogosyan, S. Prunet, A. C. S. Readhead, M. C. Shepherd, P. S. Udomprasert, L. Bronfman, W. L. Holzapfel, and J. May. Cosmological parameters from Cosmic Background Imager observations and comparisons with BOOMERanG, DASI, and MAXIMA. May 2002, arXiv:astro-ph/0205387.

[83] Davison E. Soper. *Classical Field Theory.* John Wiley & Sons, Ltd., 1976.

[84] D. N. Spergel, L. Verde, H. V. Peiris, E. Komatsu, M. R. Nolta, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Determination of cosmological parameters. *The Astrophysical Journal Supplement Series*, 148(1):175–194, September 2003, arXiv:astro-ph/0302209.

[85] Alexei A. Starobinsky. A new type of isotropic cosmological models without singularity. *Physics Letters*, B91(1):99–102, 1980.

[86] Alexei A. Starobinsky. Future and origin of our universe: Modern view. *Gravitation and Cosmology*, 6(3):157–163, September 2000, arXiv:astro-ph/9912054.

[87] Ewan D. Stewart and David H. Lyth. A more accurate analytic calculation of the spectrum of cosmological perturbations produced during inflation. *Physics Letters*, B302(2–3):171–175, March 1993, arXiv:gr-qc/9302019.

[88] John R. Taylor. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements.* University Science Books, 1982.

[89] Richard Chase Tolman. Effect of inhomogeneity on cosmological models. *Proceedings of the National Academy of Sciences, USA*, 20(3):169–176, 1934. Reprinted in [90].

[90] Richard Chase Tolman. Effect of inhomogeneity on cosmological models. *General Relativity and Gravitation*, 29(7):935–943, July 1997.

[91] Masayuki Umemura, Abraham Loeb, and Edwin L. Turner. Early cosmic formation of massive black holes. *Astrophysical Journal Letters*, 419(2):459–468, December 1993, arXiv:astro-ph/9303004.

[92] N. G. van Kampen. *Stochastic Processes in Physics and Chemistry.* North-Holland Publishing Company, 1981.

[93] Erik Vanmarcke. *Random Fields: Analysis and Synthesis.* The MIT Press, 1983.

[94] Alexander Vilenkin. Quantum fluctuations in the new inflationary universe. *Nuclear Physics*, B226(2):527–546, 1983.

[95] Steven Weinberg. Anthropic bound on the cosmological constant. *Physical Review Letters*, 59(22):2607–2610, November 1987.

[96] Steven Weinberg. *The Quantum Theory of Fields*, volume 1. Cambridge University Press, 1995.

[97] Steven Weinberg. Theories of the cosmological constant. October 1996, arXiv:astro-ph/9610044.

[98] Wolfram Research, Inc. A comprehensive online compendium of formulas involving the special functions of mathematics. `http://functions.wolfram.com`, 2001. A particular equation can be retrieved by appending it to the URL, eg. `http://functions.wolfram.com/07.25.03.0006.01`.

[99] Ya. B. Zeldovich. A hypothesis, unifying the structure and the entropy of the universe. *Monthly Notices of the Royal Astronomical Society, Short Communication*, 160:1–4, 1972.

Appendix A

Units and Sign Conventions

For detailed values of the physical constants used, see Appendix B. In terms of $\hbar$, c, and G the definitions of the Planck dimensions are[1]

$$l_{\mathrm{Pl}} = \sqrt{\frac{\hbar G}{c^3}} \qquad t_{\mathrm{Pl}} = \sqrt{\frac{\hbar G}{c^5}} \qquad m_{\mathrm{Pl}} = \sqrt{\frac{\hbar c}{G}} \tag{A.0.1}$$

and in terms of the Planck dimensions $\hbar$, c and G are

$$\hbar = \frac{{l_{\mathrm{Pl}}}^2 m_{\mathrm{Pl}}}{t_{\mathrm{Pl}}} \qquad c = \frac{l_{\mathrm{Pl}}}{t_{\mathrm{Pl}}} \qquad G = \frac{{l_{\mathrm{Pl}}}^3}{m_{\mathrm{Pl}} {t_{\mathrm{Pl}}}^2}. \tag{A.0.2}$$

The Planck system of units is is defined by setting $l_{\mathrm{Pl}} = t_{\mathrm{Pl}} = m_{\mathrm{Pl}} = 1$ which is equivalent to saying $\hbar = c = G = 1$. This document uses Planck units.

The sign convention used throughout this document is that of Misner, Thorne and Wheeler which makes the choices of sign indicated by the arrows below.

$$\overset{\downarrow}{+}\, \mathrm{d}s^2 = -\,\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2$$

$$\overset{\downarrow}{+}\, R^{\mu}{}_{\nu\alpha\beta} = \Gamma^{\mu}{}_{\nu\beta,\alpha} - \Gamma^{\mu}{}_{\nu\alpha,\beta} + \Gamma^{\mu}{}_{\sigma\alpha}\Gamma^{\sigma}{}_{\nu\beta} - \Gamma^{\mu}{}_{\sigma\beta}\Gamma^{\sigma}{}_{\nu\alpha}$$

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \overset{\downarrow}{+}\, 8\pi T_{\mu\nu}$$

This choice of sign is also called the Landau-Lifshitz space-like convention and is described as $(+++)$ in the notation of Misner, Thorne and Wheeler. In this notation, the three $+$ signs indicate, in order, the choices of sign indicated above. The inside front cover of [65] has a table of the conventions used by some other authors.

---

[1] People sometimes define the "reduced" Planck dimensions by using $8\pi G$ wherever G is used here. See, for example, [55, page 78] where $m_{\mathrm{Pl}} = \sqrt{\hbar c/(8\pi G)}$. People use the same symbols for these quantities and sometimes refer to them as simply the "Planck dimensions" which results in confusion (contrast this to the case of h and $\hbar$ which differ by a factor of 2π, are also often called by the same name, but are *not* given the same symbol). The motivation for this alternate definition is that it removes the 8π from Einstein's equation. The definitions used in this document, which do not include the 8π, are those specified by the International Council for Science, Committee on Data for Science and Technology (CODATA) [66].

Appendix B

Numerical Parameters

Some current numerical parameters are given below. Table B.1 contains physical constants and conversion factors taken from [66]. Table B.2 contains conversion factors derived from those in Table B.2. Table B.3 contains additional conversion factors found in [65]. The dimensionless parametric factor h, called the Hubble parameter, that appears in some of the values in Table B.5 is used to encode the uncertainty in the Hubble constant and indicates how that uncertainty propagates to other quantities. Its use is not exhaustive however: some quantities that rely on the value of h do not have that dependence explicitly indicated.

speed of light in vacuum	c	=	299792458 m/s
Newton constant of gravitation	G	=	$6.673(10) \times 10^{-11}$ m^3/kg/s^2
Planck constant	$\hbar$	=	$1.054571596(82) \times 10^{-34}$ Js
		=	$6.58211889(26) \times 10^{-16}$ eVs
Planck mass	m_{Pl}	=	$2.1767(16) \times 10^{-8}$ kg
Planck length	l_{Pl}	=	$1.6160(12) \times 10^{-35}$ m
Planck time	t_{Pl}	=	$5.3906(40) \times 10^{-44}$ s
Boltzmann constant	k_{B}	=	$1.3806503(24) \times 10^{-23}$ J/K
		=	$8.617342(15) \times 10^{-5}$ eV/K
electron volt	1 eV	=	$1.602176462(63) \times 10^{-19}$ J

Table B.1: Physical constants and conversion factors. Taken from [66].

Planck density	ρ_{Pl}	$= \frac{m_{\mathrm{Pl}}}{{l_{\mathrm{Pl}}}^3} = \frac{c^5}{\hbar G^2}$
		$= 5.1579 \times 10^{96}$ kg/m^3
Planck energy	E_{Pl}	$= m_{\mathrm{Pl}} c^2 = \sqrt{\frac{\hbar c^5}{G}}$
		$= 1.9563 \times 10^{9}$ J
		$= 1.2210 \times 10^{28}$ eV
Boltzmann constant	k_{B}	$= 7.0575 \times 10^{-33}$ K^{-1}
Planck temperature	T_{Pl}	$= \frac{1}{k_{\mathrm{B}}}$
		$= 1.4169 \times 10^{32}$ K

Table B.2: Derived physical constants and conversion factors.

Distance	1 pc	$= 3.0856 \times 10^{16}$ m
		$= 1.9094 \times 10^{51}\ l_{\mathrm{Pl}}$
	1 light year	$= 9.4605 \times 10^{15}$ m
		$= 5.8542 \times 10^{50}\ l_{\mathrm{Pl}}$
	1 AU	$= 1.495985 \times 10^{11}$ m
		$= 9.2574 \times 10^{45}\ l_{\mathrm{Pl}}$
Time	1 a	= 31556926 s
		$= 5.8542 \times 10^{50}\ t_{\mathrm{Pl}}$
	1 d	= 86400 s
		$= 1.6028 \times 10^{48}\ t_{\mathrm{Pl}}$
	1 sidereal day	= 86164.091 s
		$= 1.5984 \times 10^{48}\ t_{\mathrm{Pl}}$
Mass, Energy, Temperature	1 eV	$= 1.78268 \times 10^{-36}$ kg
		$= 1.324 \times 10^{-63}$ m

Table B.3: Additional conversion factors. Partially taken from the back cover of [65].

Hubble radius	1 H_0^{-1}	$= 2.99 h^{-1}$ Gpc
Hubble time	1 H_0^{-1}	$= 9.7 h^{-1}$ Ga

Table B.4: Conversion factors derived from cosmological parameters. See Table B.5.

Mass of Sun	$M_\odot = 1.989 \times 10^{30}$ kg $= 9.138 \times 10^{37}\ m_{\mathrm{Pl}}$
Mass of Earth	$M_\oplus = 5.977 \times 10^{24}$ kg $= 2.746 \times 10^{32}\ m_{\mathrm{Pl}}$
Mass of our galaxy	$M_{\mathrm{G}} = 1.8 \times 10^{11}\ \mathrm{M}_\odot = 1.6 \times 10^{49}\ m_{\mathrm{Pl}}$
Diameter of our galaxy	$D_{\mathrm{G}} = 1 \sim 1.5 \times 10^5$ light year $= 6 \sim 9 \times 10^{55}\ l_{\mathrm{Pl}}$
Our local group	$M = 16$ members $D \approx 0.4$ Mpc $= 8 \times 10^{56}\ l_{\mathrm{Pl}}$
Virgo cluster	$M = 2500$ members $D \approx 12$ Mpc $= 2.3 \times 10^{58}\ l_{\mathrm{Pl}}$
Typical cluster	$M = 130$ members $D \approx 3$ Mpc $= 6 \times 10^{57}\ l_{\mathrm{Pl}}$
Cosmological age [82]	$t_0 = 13.7 \pm 0.2$ Ga
Age at decoupling [10]	$t_{\mathrm{dec}} = 379^{+8}_{-7}$ ka
Smallest scale that is still linear today [55, Section 4.2.5]	$D_{\mathrm{linear}} = 8h^{-1}$ Mpc $= 2h^{-1} \times 10^{58}\ l_{\mathrm{Pl}}$
Hubble parameter [82]	$h = 0.69 \pm 0.02$
Hubble constant	$H_0 = 100h$ km/s/Mpc $= 1.75h \times 10^{-61}\ {t_{\mathrm{Pl}}}^{-1}$
Critical density	$\rho_{\mathrm{c}} = 1.88 \times 10^{-26} h^2\ \mathrm{kg/m^3} = 3.64 \times 10^{-123} h^2\ \rho_{\mathrm{Pl}}$
Density parameter [64, Figure 7]	$0.95 \leq \Omega_0 \leq 1.18$ (95% confidence interval)
Dark matter contribution [82]	$\Omega_{\mathrm{m}} h^2 = 0.12^{+0.03}_{-0.03}$
Baryon contribution [82]	$\Omega_{\mathrm{b}} h^2 = 0.022^{+0.015}_{-0.009}$
Dark energy contribution [82]	$\Omega_\Lambda = 0.70^{+0.02}_{-0.03}$
Dark energy equation of state [10]	$w_\Lambda < -0.78$ (95% confidence level)
Primeval spectral index [82]	$n_{\mathrm{s}} = 1.01^{+0.09}_{-0.08}$
Matter density contrast at horizon entry [55, Section 9.1.2]	$\delta_{\mathrm{H}}(k) = 1.94 \times 10^{-5} \pm 7\%$ (1-sigma confidence interval)
Cosmic background temperature [24]	$T_\gamma = 2.728 \pm 0.004\ \mathrm{K} = (1.925 \pm .003) \times 10^{-32}$
Temperature at time of last photon scattering [72, equation (2.61)]	$T_{\mathrm{eq}} = 9.24 \Omega h^2\ \mathrm{eV} = 1.07 \Omega h^2 \times 10^5\ \mathrm{K} = 7.57 \Omega h^2 \times 10^{-28}$

Table B.5: Cosmological parameters.

Appendix C

Derivations

C.1 Solving Einstein's Field Equation for the Geometry of Space-Time

This appendix presents the technique used to obtain solutions to Einstein's equation used throughout this document. The procedure is:

- A line element, or guess, for the geometry is proposed along with an expression for the stress-energy tensor.
- The Einstein tensor is computed from the line element.
- The Einstein tensor and stress-energy tensor are substituted into Einstein's equation, and the equations of motion fall out.

This procedure may not the most efficient technique for doing this but it is the most mechanical, requiring the least amount of thought. For this reason it is also the simplest technique and, in principle, can be used for any problem.

The information needed to accomplish this task can be found in the following definitions. The components of the Einstein tensor are given by [65, page 224]

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R. \tag{C.1.1}$$

The curvature scalar, R, is given by [65, page 224]

$$R \equiv g^{\mu\nu} R_{\mu\nu}. \tag{C.1.2}$$

The components of the Ricci tensor, $R_{\mu\nu}$, are given by [65, page 224]

$$R_{\mu\nu} \equiv R^{\alpha}{}_{\mu\alpha\nu} = \Gamma^{\alpha}{}_{\mu\nu,\alpha} - \Gamma^{\alpha}{}_{\mu\alpha,\nu} + \Gamma^{\alpha}{}_{\beta\alpha}\Gamma^{\beta}{}_{\mu\nu} - \Gamma^{\alpha}{}_{\beta\mu}\Gamma^{\beta}{}_{\nu\alpha} \tag{C.1.3}$$

where ${R^\alpha}_{\beta\gamma\delta}$ are the components of the Riemann curvature tensor. The Christoffel symbols, ${\Gamma^\alpha}_{\beta\gamma}$, are given by [65, page 223]

$$ {\Gamma^\alpha}_{\beta\gamma} = \frac{1}{2} g^{\alpha\delta} \left(g_{\delta\beta\,,\gamma} + g_{\delta\gamma\,,\beta} - g_{\beta\gamma\,,\delta} \right) . \tag{C.1.4} $$

With this information, one can see the procedure to be followed. It is summarized in the following flow charts.

$$ \begin{gathered} g_{\mu\nu} \to {\Gamma^\alpha}_{\beta\gamma} \to R_{\mu\nu} \to R \\ g_{\mu\nu} \oplus R_{\mu\nu} \oplus R \to G_{\mu\nu} \\ G_{\mu\nu} \oplus T_{\mu\nu} \to \text{equations of motion} \end{gathered} $$

Calculating all 64 Christoffel symbols from the definition above, even skipping the 24 duplicates, is extremely tedious. Luckily it can be shown that for diagonal metrics, the only non-zero components are those with repeated indices. These are then given by the simpler expressions

$$ {\Gamma^\alpha}_{\beta\beta} = -\frac{1}{2} \frac{1}{g_{\alpha\alpha}} g_{\beta\beta\,,\alpha} \qquad , \alpha \neq \beta \tag{C.1.5a} $$

$$ {\Gamma^\alpha}_{\beta\alpha} = {\Gamma^\alpha}_{\alpha\beta} = \frac{1}{2} \frac{1}{g_{\alpha\alpha}} g_{\alpha\alpha\,,\beta} \qquad , \alpha \neq \beta \tag{C.1.5b} $$

$$ {\Gamma^\alpha}_{\alpha\alpha} = \frac{1}{2} \frac{1}{g_{\alpha\alpha}} g_{\alpha\alpha\,,\alpha} \tag{C.1.5c} $$

where the summation convention is suppressed. See [57, Problem 7.6].

C.2 Christoffel Symbols for the Robertson-Walker Line Element

The covariant components of the metric can be read directly off of (2.1.3) and are

$$ \begin{array}{llll} g_{tt} = -1 & g_{tr} = 0 & g_{t\theta} = 0 & g_{t\phi} = 0 \\ g_{rt} = 0 & g_{rr} = \frac{a^2}{1-kr^2} & g_{r\theta} = 0 & g_{r\phi} = 0 \\ g_{\theta t} = 0 & g_{\theta r} = 0 & g_{\theta\theta} = a^2 r^2 & g_{\theta\phi} = 0 \\ g_{\phi t} = 0 & g_{\phi r} = 0 & g_{\phi\theta} = 0 & g_{\phi\phi} = a^2 r^2 \sin^2\theta \end{array} \tag{C.2.1} $$

From the components of the metric tensor and the special-case expressions for the Christoffel symbols in (C.1.5), the Christoffel symbols for this line element are

$$\begin{array}{l|l|l|l}
\Gamma^{t}{}_{tt}=0 & \Gamma^{r}{}_{tt}=0 & \Gamma^{\theta}{}_{tt}=0 & \Gamma^{\phi}{}_{tt}=0 \\
\Gamma^{t}{}_{tr}=0 & \Gamma^{r}{}_{tr}=\frac{\dot{a}}{a} & \Gamma^{\theta}{}_{tr}=0 & \Gamma^{\phi}{}_{tr}=0 \\
\Gamma^{t}{}_{t\theta}=0 & \Gamma^{r}{}_{t\theta}=0 & \Gamma^{\theta}{}_{t\theta}=\frac{\dot{a}}{a} & \Gamma^{\phi}{}_{t\theta}=0 \\
\Gamma^{t}{}_{t\phi}=0 & \Gamma^{r}{}_{t\phi}=0 & \Gamma^{\theta}{}_{t\phi}=0 & \Gamma^{\phi}{}_{t\phi}=\frac{\dot{a}}{a} \\
\hline
\Gamma^{t}{}_{rt}=0 & \Gamma^{r}{}_{rt}=\frac{\dot{a}}{a} & \Gamma^{\theta}{}_{rt}=0 & \Gamma^{\phi}{}_{rt}=0 \\
\Gamma^{t}{}_{rr}=\frac{a\dot{a}}{1-k^2} & \Gamma^{r}{}_{rr}=\frac{kr}{1-kr^2} & \Gamma^{\theta}{}_{rr}=0 & \Gamma^{\phi}{}_{rr}=0 \\
\Gamma^{t}{}_{r\theta}=0 & \Gamma^{r}{}_{r\theta}=0 & \Gamma^{\theta}{}_{r\theta}=\frac{1}{r} & \Gamma^{\phi}{}_{r\theta}=0 \\
\Gamma^{t}{}_{r\phi}=0 & \Gamma^{r}{}_{r\phi}=0 & \Gamma^{\theta}{}_{r\phi}=0 & \Gamma^{\phi}{}_{r\phi}=\frac{1}{r} \\
\hline
\Gamma^{t}{}_{\theta t}=0 & \Gamma^{r}{}_{\theta t}=0 & \Gamma^{\theta}{}_{\theta t}=\frac{\dot{a}}{a} & \Gamma^{\phi}{}_{\theta t}=0 \\
\Gamma^{t}{}_{\theta r}=0 & \Gamma^{r}{}_{\theta r}=0 & \Gamma^{\theta}{}_{\theta r}=\frac{1}{r} & \Gamma^{\phi}{}_{\theta r}=0 \\
\Gamma^{t}{}_{\theta\theta}=a\dot{a}r^2 & \Gamma^{r}{}_{\theta\theta}=-(1-kr^2)r & \Gamma^{\theta}{}_{\theta\theta}=0 & \Gamma^{\phi}{}_{\theta\theta}=0 \\
\Gamma^{t}{}_{\theta\phi}=0 & \Gamma^{r}{}_{\theta\phi}=0 & \Gamma^{\theta}{}_{\theta\phi}=0 & \Gamma^{\phi}{}_{\theta\phi}=\frac{\cos\theta}{\sin\theta} \\
\hline
\Gamma^{t}{}_{\phi t}=0 & \Gamma^{r}{}_{\phi t}=0 & \Gamma^{\theta}{}_{\phi t}=0 & \Gamma^{\phi}{}_{\phi t}=\frac{\dot{a}}{a} \\
\Gamma^{t}{}_{\phi r}=0 & \Gamma^{r}{}_{\phi r}=0 & \Gamma^{\theta}{}_{\phi r}=0 & \Gamma^{\phi}{}_{\phi r}=\frac{1}{r} \\
\Gamma^{t}{}_{\phi\theta}=0 & \Gamma^{r}{}_{\phi\theta}=0 & \Gamma^{\theta}{}_{\phi\theta}=0 & \Gamma^{\phi}{}_{\phi\theta}=\frac{\cos\theta}{\sin\theta} \\
\Gamma^{t}{}_{\phi\phi}=a\dot{a}r^2\sin^2\theta & \Gamma^{r}{}_{\phi\phi}=-(1-kr^2)r\sin^2\theta & \Gamma^{\theta}{}_{\phi\phi}=-\sin\theta\cos\theta & \Gamma^{\phi}{}_{\phi\phi}=0
\end{array} \tag{C.2.2}$$

where a dot indicates differentiation with respect to t.

C.3 Friedman-Robertson-Walker Mass Continuity Equation in (2.1.7)

Differentiating (2.1.6b) with respect to time gives

$$2\left(\frac{\dot{a}}{a}\right)\left(\frac{\ddot{a}a-\dot{a}^2}{a^2}\right)-2\frac{k}{a^3}\dot{a}=\frac{8\pi}{3}\dot{\rho} \tag{C.3.1}$$

or

$$2\left(\frac{\ddot{a}}{a}\right)-2\left(\frac{\dot{a}}{a}\right)^2-2\frac{k}{a^2}=\frac{8\pi}{3}\dot{\rho}\left(\frac{a}{\dot{a}}\right). \tag{C.3.2}$$

Adding 3× (2.1.6b) gives

$$2\left(\frac{\ddot{a}}{a}\right)+\left(\frac{\dot{a}}{a}\right)^2+\frac{k}{a^2}=\frac{8\pi}{3}\dot{\rho}\left(\frac{a}{\dot{a}}\right)+8\pi\rho. \tag{C.3.3}$$

Subtracting (2.1.6a) leaves

$$-8\pi p=\frac{8\pi}{3}\dot{\rho}\left(\frac{a}{\dot{a}}\right)+8\pi\rho \tag{C.3.4}$$

or

$$\dot{\rho}a^3+3\left(\rho+p\right)a^2\dot{a}=0, \tag{C.3.5}$$

the energy conservation equation.[1] This can also be written as[2]

$$\frac{\dot{\rho}}{\rho+p} = -3\frac{\dot{a}}{a}. \tag{C.3.6}$$

C.4 Equation (2.1.9)

Substituting $p = w\rho$ into (2.1.7) gives

$$\dot{\rho}a^3 + 3\left(\rho + w\rho\right)a^2\dot{a} = 0 \tag{C.4.1}$$

therefore

$$\frac{1}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}t} = -3\left(1+w\right)\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}t} \tag{C.4.2}$$

or

$$\frac{\mathrm{d}\rho}{\rho} = -3\left(1+w\right)\frac{\mathrm{d}a}{a}. \tag{C.4.3}$$

Integrating gives

$$\ln\rho = -3\left(1+w\right)\ln a \tag{C.4.4}$$

to within an additive constant, so

$$\rho \propto a^{-3(1+w)}. \tag{C.4.5}$$

C.5 Equation of Motion for the Scalar Field in (2.2.1)

The Euler-Lagrange equation is[3]

$$\left(\frac{\partial L}{\partial\phi_{;\mu}}\right)_{;\mu} - \frac{\partial L}{\partial\phi} = 0 \tag{C.5.1}$$

and from (2.2.1)

$$L = -\frac{1}{2}g^{\rho\sigma}\phi_{;\rho}\phi_{;\sigma} - \frac{1}{2}\left(m^2 + \xi R\right)\phi^2. \tag{C.5.2}$$

So we get

$$\frac{\partial L}{\partial\phi_{;\mu}} = -\frac{1}{2}g^{\rho\sigma}\phi_{;\rho}\delta^{\mu}{}_{\sigma} - \frac{1}{2}g^{\rho\sigma}\phi_{;\sigma}\delta^{\mu}{}_{\rho} = -\phi^{;\mu}$$

and

$$\frac{\partial L}{\partial\phi} = -\left(m^2 + \xi R\right)\phi.$$

---

[1] This agrees with [60, equation 1.3.8].

[2] This agrees with [72, equation (4.40)].

[3] See [83, equation (3.14)].

The equation of motion for the field is, therefore,

$$\phi^{;\mu}{}_{;\mu} = \Box\phi = \left(m^2 + \xi R\right)\phi = \frac{\mathrm{d}V}{\mathrm{d}\phi}. \tag{C.5.3}$$

The time and spatial parts of the d'Alembertian can be isolated as follows. For any diagonal metric

$$\begin{aligned}
\phi^{;\mu}{}_{;\mu} &= \left(g^{\mu\nu}\phi_{;\nu}\right)_{;\mu} = \left(g^{\mu\nu}\phi_{,\nu}\right)_{;\mu} = g^{\mu\nu}\left(\phi_{,\nu}\right)_{;\mu} \\
&= g^{\mu\nu}\left(\phi_{,\nu\mu} - \Gamma^{\alpha}{}_{\mu\nu}\phi_{,\alpha}\right) \\
\phi^{;\mu}{}_{;\mu} &= g^{00}\left(\phi_{,00} - \Gamma^{\alpha}{}_{00}\phi_{,\alpha}\right) - g^{ii}\Gamma^{0}{}_{ii}\phi_{,0} + g^{ij}\left(\phi_{,ij} - \Gamma^{\alpha}{}_{ij}\phi_{,\alpha}\right).
\end{aligned} \tag{C.5.4}$$

Comparing the last term on the last line with the second line one sees that if one factors $\frac{1}{a^2}$ out of the g^{ij} then one is left with an expression which is equivalent to the covariant Laplacian on the unscaled spatial part of the metric so

$$\phi^{;\mu}{}_{;\mu} = g^{00}\left(\phi_{,00} - \Gamma^{\alpha}{}_{00}\phi_{,\alpha}\right) - g^{ii}\Gamma^{0}{}_{ii}\phi_{,0} + \frac{1}{a^2}\Delta\phi$$

where $|$ denotes the covariant derivative with respect to the unscaled spatial metric and $\Delta\phi = \phi^{|i}{}_{|i}$. Inserting the expressions for the inverse metric components and the Christoffel symbols from (C.2.1) and (C.2.2) gives

$$\begin{aligned}
\phi^{;\mu}{}_{;\mu} &= -\ddot{\phi} - \left(g^{rr}\Gamma^{t}{}_{rr} + g^{\theta\theta}\Gamma^{t}{}_{\theta\theta} + g^{\phi\phi}\Gamma^{t}{}_{\phi\phi}\right)\dot{\phi} + \frac{1}{a^2}\Delta\phi \\
&= -\ddot{\phi} - 3\frac{\dot{a}}{a}\dot{\phi} + \frac{1}{a^2}\Delta\phi.
\end{aligned}$$

Substituting this into (C.5.3) gives[4]

$$\ddot{\phi} + 3\frac{\dot{a}}{a}\dot{\phi} - \frac{1}{a^2}\Delta\phi = -\frac{\mathrm{d}V}{\mathrm{d}\phi} \tag{C.5.5}$$

where Δ is the covariant Laplacian on the metric defined by

$$\mathrm{d}l^2 = \frac{\mathrm{d}r^2}{1 - kr^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right), \tag{C.5.6}$$

i.e. the unscaled spatial part of the Robertson-Walker metric. Making the assumption that the inflaton field is sufficiently smooth that its energy density is dominated by its potential rather than its spatial derivatives, (C.5.5) is simplified to just

$$\ddot{\phi} + 3\frac{\dot{a}}{a}\dot{\phi} = -\frac{\mathrm{d}V}{\mathrm{d}\phi}. \tag{C.5.7}$$

---

[4] Compare this result to that in [60, page 44] but note that Linde is using a different sign convention for $g_{\mu\nu}$ which leads to a sign difference in $V(\phi)$.

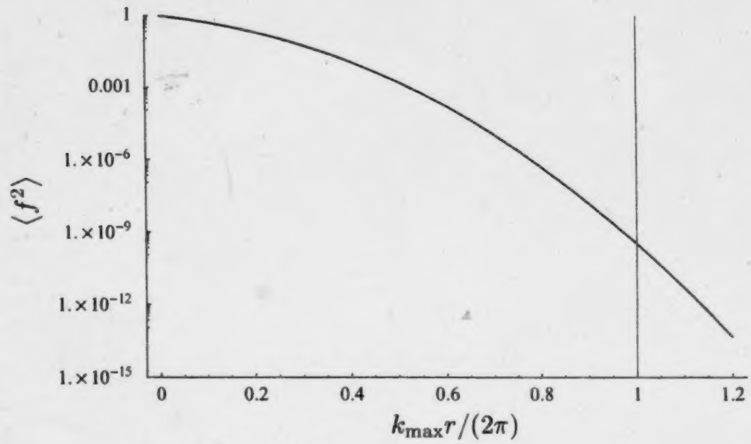


Figure C.1: The result of the integral in (C.6.1) as a function of the upper bound, $k_{\max}$, for a Gaussian window. The reciprocal smoothing radius is indicated at $k_{\max} = 2\pi/r$. The vertical axis is the fractional error when compared to the exact result given by $k_{\max} = \infty$.

C.6 Justification for the Removal of the Phase Factor in (3.4.40)

Since it is difficult to set up tests related to the actual phase factor in question, a more simple approach will be taken. We will compute the variance for an unspecified homogeneous isotropic Gaussian random field with a mean of 0 and a $n_s = 1$ that has been filtered by the window function $W_{\tilde{k};r}$. We will then examine the effect of truncating the integral at various wave-numbers. This is a straight-forward procedure that has a significant effect on high frequency modes and will establish whether modifications to them influence the results in any great way. To be precise, we will compute

$$\langle f^2 \rangle = \int_0^{k_{\max}} W_{\tilde{k};r}^2 \frac{\mathrm{d}k}{k} \tag{C.6.1}$$

for various choices of $k_{\max}$. Comparison to (3.4.32) shows this to be the variance of a smoothed $n_s = 1$ field with the field's power spectrum normalized so as to make the prefactor of the integral 1. Figure C.1 shows the fractional error when compared to the exact result as a function of $k_{\max}$ when a Gaussian window function is chosen. The fractional error goes to 1 as $k_{\max}$ is taken to 0 as one would expect. As can be seen, stopping the integration at wave-numbers as small as the reciprocal smoothing radius, in other words completely throwing out all modes with wavelengths shorter than the smoothing radius, introduces a fractional error of less than 10^{-9} — a negligible effect on the result. The effect of dropping the phase factor must be expected to be insignificant.

C.7 Solution to (4.1.13)

Here I derive the solution to (4.1.13). The differential equation in question is

$$\ddot{\psi}_p(t) + 3H\dot{\psi}_p(t) + H^2p^2e^{-2Ht}\psi_p(t) = 0 \tag{C.7.1}$$

If we transform to conformal time, η, defined by

$$\eta = \int_0^t \frac{1}{a(t)}\,\mathrm{d}t = \int_0^t H\mathrm{e}^{-Ht}\,\mathrm{d}t = -\mathrm{e}^{-Ht} \tag{C.7.2}$$

we get

$$0 = \left(\frac{\mathrm{d}\eta}{\mathrm{d}t}\right)^2 \frac{\mathrm{d}^2}{\mathrm{d}\eta^2}\psi_p + \frac{\mathrm{d}\eta}{\mathrm{d}t}\left(\frac{\mathrm{d}}{\mathrm{d}\eta}\frac{\mathrm{d}\eta}{\mathrm{d}t} + 3H\right)\frac{\mathrm{d}}{\mathrm{d}\eta}\psi_p + H^2p^2\eta^2\psi_p$$

and using

$$\begin{aligned}\frac{\mathrm{d}\eta}{\mathrm{d}t} &= H\mathrm{e}^{-Ht} = -H\eta \\ \frac{\mathrm{d}}{\mathrm{d}\eta}\frac{\mathrm{d}\eta}{\mathrm{d}t} &= -H\end{aligned}$$

we get

$$0 = H^2\eta^2\frac{\mathrm{d}^2}{\mathrm{d}\eta^2}\psi_p + H^2\eta\,(1-3)\,\frac{\mathrm{d}}{\mathrm{d}\eta}\psi_p + H^2p^2\eta^2\psi_p$$

or

$$\eta^2\frac{\mathrm{d}^2}{\mathrm{d}\eta^2}\psi_p - 2\eta\frac{\mathrm{d}}{\mathrm{d}\eta}\psi_p + p^2\eta^2\psi_p = 0. \tag{C.7.3}$$

If we now let $\psi_p = \eta^s Z$ then the differential equation becomes

$$\eta^2\left[s(s-1)\eta^{s-2}Z + 2s\eta^{s-1}\frac{\mathrm{d}}{\mathrm{d}\eta}Z + \eta^s\frac{\mathrm{d}^2}{\mathrm{d}\eta^2}Z\right] - 2\eta\left[s\eta^{s-1}Z + \eta^s\frac{\mathrm{d}}{\mathrm{d}\eta}Z\right] + p^2\eta^{s+2}Z = 0$$

and dividing by η^s leaves us with

$$\eta^2\frac{\mathrm{d}^2}{\mathrm{d}\eta^2}Z + 2\,(s-1)\,\eta\frac{\mathrm{d}}{\mathrm{d}\eta}Z + \left(p^2\eta^2 + s(s-3)\right)Z = 0. \tag{C.7.4}$$

Finally, setting $s = 3/2$ we get

$$\eta^2\frac{\mathrm{d}^2}{\mathrm{d}\eta^2}Z + \eta\frac{\mathrm{d}}{\mathrm{d}\eta}Z + \left(p^2\eta^2 - \frac{9}{4}\right)Z = 0 \tag{C.7.5}$$

which is Bessel's equation of order 3/2.[5] The complete solution of this equation can be expressed as a linear combination of Hankel functions of the 1st and 2nd kinds and using these the solution for ψ_p is

---

[5] See Section C.7.1.

$$\psi_p(\eta) = C_1(p)\eta^{\frac{3}{2}}\,\mathrm{H}^{(1)}_{3/2}(p\eta) + C_2(p)\eta^{\frac{3}{2}}\,\mathrm{H}^{(2)}_{3/2}(p\eta) \tag{C.7.6}$$

where C_1 and C_2 are arbitrary constants which can depend on the parameter p.[6]

We can use the requirement that the mode functions in (4.1.10) be orthonormal under an inner product to put restrictions on C_1 and C_2. The standard definition of the inner product for scalar fields is[7]

$$(\phi_1, \phi_2) = -\mathrm{i}\int_\Sigma \phi_1(x)\,\overleftrightarrow{\partial_\mu}\,\phi_2^*(x)\sqrt{-g_\Sigma}\,\mathrm{d}\Sigma^\mu \tag{C.7.7}$$

$$= -\mathrm{i}\int_\Sigma \left[\phi_1(x)\partial_\mu\phi_2^*(x) - (\partial_\mu\phi_1(x))\,\phi_2^*(x)\right]\sqrt{-g_\Sigma}\,\mathrm{d}\Sigma^\mu \tag{C.7.8}$$

where Σ is some space-like hyper-surface over which the integral is performed, and $\sqrt{-g_\Sigma}\,\mathrm{d}\Sigma^\mu = \sqrt{-g_\Sigma}\,n^\mu\,\mathrm{d}\Sigma$ where n^μ is the future-directed unit vector orthogonal to the hyper-surface and $\sqrt{-g_\Sigma}\,\mathrm{d}\Sigma$ is the proper volume element in the surface. The orthonormality requirements on the mode functions are[8]

$$\left(u_p, u_{p'}\right) = \delta^3\left(\vec{p} - \vec{p}'\right) \tag{C.7.9}$$

$$\left(u_p^*, u_{p'}^*\right) = -\delta^3\left(\vec{p} - \vec{p}'\right) \tag{C.7.10}$$

$$\left(u_p, u_{p'}^*\right) = 0 \tag{C.7.11}$$

Because of the choice of variable separation made when writing the mode functions in (4.1.10), the easiest space-like hyper-surface to choose for an inner product is the one perpendicular to the t-direction so

$$n^\mu = \delta^\mu{}_t$$

and

$$\sqrt{-g_\Sigma}\,\mathrm{d}\Sigma = a^3\,\mathrm{d}^3x.$$

With this, (C.7.9) gives

$$\begin{aligned}
\delta^3\left(\vec{p} - \vec{p}'\right) &= -\mathrm{i}\int a^3\,\mathrm{d}^3x\,\left[u_p(x)\partial_t u_{p'}^*(x) - (\partial_t u_p(x))\,u_{p'}^*(x)\right] \\
&= -\mathrm{i}\left(\frac{a}{2\pi}\right)^3\int \mathrm{d}^3x\,\left[\psi_p(t)\mathrm{e}^{\mathrm{i}\vec{p}\cdot\vec{x}}\mathrm{e}^{-\mathrm{i}\vec{p}'\cdot\vec{x}}\partial_t\psi_{p'}^*(t) - \psi_{p'}^*(t)\mathrm{e}^{-\mathrm{i}\vec{p}'\cdot\vec{x}}\mathrm{e}^{\mathrm{i}\vec{p}\cdot\vec{x}}\partial_t\psi_p(t)\right]
\end{aligned}$$

---

[6] Yes, the *H*'s are confusing: the Hankel functions can be distinguished from the Hubble constant by the superscripted numbers in parentheses indicating the kind of Hankel function. The Hankel functions, being operators, are also typeset as an upright H while the Hubble constant, being a variable, is typeset as an italic *H*.

[7] See [12, equation (3.28)].

[8] See [12, equation (3.29)]. The integrand in the inner product is the Klein-Gordon four-current for the field (see [34, page 260]). Projecting this onto the surface's unit normal vector picks out the field's probability density in the rest frame of an observer whose four-velocity is normal to the surface. We are requiring the integral of this on a three-surface to be 1 at $p = p'$.

$$= -\mathrm{i}a^3\left[\psi_p(t)\partial_t\psi_{p'}^*(t) - \psi_{p'}^*(t)\partial_t\psi_p(t)\right]\frac{1}{(2\pi)^3}\int \mathrm{e}^{\mathrm{i}(\vec{p}-\vec{p}')\cdot\vec{x}}\,\mathrm{d}^3x.$$

The integral of the exponential along with the $(2\pi)^{-3}$ is equal to $\delta^3(\vec{p}-\vec{p}')$ so as long as the rest of the stuff on the right hand side equals 1 when $\vec{p}=\vec{p}'$ this orthonormality condition will be satisfied. Continuing with the normalization condition we find

$$\begin{aligned}
\frac{\mathrm{i}}{a^3} &= \psi_p(t)\partial_t\psi_p^*(t) - \psi_p^*(t)\partial_t\psi_p(t)\\
\left(\frac{\mathrm{d}\eta}{\mathrm{d}t}\right)^{-1}\frac{\mathrm{i}}{a^3} &= \psi_p(\eta)\partial_\eta\psi_p^*(\eta) - \psi_p^*(\eta)\partial_\eta\psi_p(\eta)\\
\frac{-\mathrm{i}}{a^3H\eta} &= \psi_p(\eta)\partial_\eta\psi_p^*(\eta) - \psi_p^*(\eta)\partial_\eta\psi_p(\eta)\\
\frac{-\mathrm{i}}{a^3H\eta^4} &= |C_1|^2\left[\mathrm{H}_{3/2}^{(1)}(p\eta)\partial_\eta\,\mathrm{H}_{3/2}^{(1)*}(p\eta) - \mathrm{H}_{3/2}^{(1)*}(p\eta)\partial_\eta\,\mathrm{H}_{3/2}^{(1)}(p\eta)\right] +\\
&\qquad |C_2|^2\left[\mathrm{H}_{3/2}^{(2)}(p\eta)\partial_\eta\,\mathrm{H}_{3/2}^{(2)*}(p\eta) - \mathrm{H}_{3/2}^{(2)*}(p\eta)\partial_\eta\,\mathrm{H}_{3/2}^{(2)}(p\eta)\right] +\\
&\qquad C_1C_2^*\left[\mathrm{H}_{3/2}^{(1)}(p\eta)\partial_\eta\,\mathrm{H}_{3/2}^{(2)*}(p\eta) - \mathrm{H}_{3/2}^{(2)*}(p\eta)\partial_\eta\,\mathrm{H}_{3/2}^{(1)}(p\eta)\right] +\\
&\qquad C_1^*C_2\left[\mathrm{H}_{3/2}^{(2)}(p\eta)\partial_\eta\,\mathrm{H}_{3/2}^{(1)*}(p\eta) - \mathrm{H}_{3/2}^{(1)*}(p\eta)\partial_\eta\,\mathrm{H}_{3/2}^{(2)}(p\eta)\right]
\end{aligned}$$

$$\frac{-\mathrm{i}}{pa^3H\eta^4} = \left[|C_1|^2 - |C_2|^2\right]\left[\mathrm{H}_{3/2}^{(1)}(p\eta)\frac{d}{d(p\eta)}\mathrm{H}_{3/2}^{(2)}(p\eta) - \mathrm{H}_{3/2}^{(1)*}(p\eta)\frac{d}{d(p\eta)}\mathrm{H}_{3/2}^{(2)*}(p\eta)\right].$$

At this point, by tracing the $*$'s through the calculation, it's easy to see that (C.7.10) also gives this condition and that (C.7.11) will be automatically satisfied simply by the form of the mode functions. Now

$$\mathrm{H}_{3/2}^{(2)}(x) = -\sqrt{\frac{2}{\pi x}}\mathrm{e}^{\mathrm{i}x}\left(1-\frac{1}{\mathrm{i}x}\right)$$

so

$$\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{H}_{3/2}^{(2)}(x) = -\mathrm{i}\sqrt{\frac{2}{\pi x}}\mathrm{e}^{\mathrm{i}x}\left(1-\frac{1}{\mathrm{i}x}-\frac{1}{x^2}\right)$$

and so

$$\begin{aligned}
\mathrm{H}_{3/2}^{(1)}(x)\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{H}_{3/2}^{(2)}(x) &= \left[-\sqrt{\frac{2}{\pi x}}\mathrm{e}^{-\mathrm{i}x}\left(1+\frac{1}{\mathrm{i}x}\right)\right]\left[-\mathrm{i}\sqrt{\frac{2}{\pi x}}\mathrm{e}^{\mathrm{i}x}\left(1-\frac{1}{\mathrm{i}x}-\frac{1}{x^2}\right)\right]\\
&= \frac{2}{\pi x}\left(\mathrm{i}-\frac{1}{x^3}\right),
\end{aligned}$$

therefore

$$\mathrm{H}_{3/2}^{(1)}(x)\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{H}_{3/2}^{(2)}(x) - \mathrm{H}_{3/2}^{(1)*}(x)\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{H}_{3/2}^{(2)*}(x) = \frac{4\mathrm{i}}{\pi x}.$$

With this,

$$|C_1|^2 - |C_2|^2 = -\frac{\pi}{4a^3 H \eta^3}$$

but

$$\eta = -\frac{1}{Ha}$$

so the normalization condition becomes

$$|C_1(p)|^2 - |C_2(p)|^2 = \frac{\pi H^2}{4}. \tag{C.7.12}$$

For convenience, we can take this factor out of the constants by doing

$$C_i(p) \longrightarrow \frac{\sqrt{\pi}}{2} H c_i(p)$$

The final, normalized, solution to the equation of motion is therefore

$$\psi_p(\eta) = \frac{\sqrt{\pi}}{2} H \eta^{\frac{3}{2}} \left[c_1(p)\, \mathrm{H}^{(1)}_{3/2}(p\eta) + c_2(p)\, \mathrm{H}^{(2)}_{3/2}(p\eta) \right] \tag{C.7.13}$$

where $|c_1(p)|^2 - |c_2(p)|^2 = 1$.[9]

Further restrictions can be placed on $c_1(p)$ and $c_2(p)$ by requiring the high frequency limit, $p \to \infty$, of the solution to match the solution obtained in Minkowski space-time which is

$$\psi_p(t) = \frac{1}{\sqrt{2p}} \mathrm{e}^{-\mathrm{i}pt}. \tag{C.7.14}$$

Starting by substituting explicit expressions for the Hankel functions into (C.7.13) we obtain

$$\begin{aligned}
\psi_p(\eta) &= -\frac{1}{\sqrt{2p}} H\eta \left[c_1(p) \left(1 + \frac{1}{\mathrm{i}p\eta}\right) \mathrm{e}^{-\mathrm{i}p\eta} + c_2(p) \left(1 - \frac{1}{\mathrm{i}p\eta}\right) \mathrm{e}^{\mathrm{i}p\eta} \right] \\
&\approx -\frac{1}{\sqrt{2p}} H\eta \left[c_1(p)\mathrm{e}^{-\mathrm{i}p\eta} + c_2(p)\mathrm{e}^{\mathrm{i}p\eta} \right] \\
&= \frac{1}{a\sqrt{2p}} \left[c_1(p)\mathrm{e}^{-\mathrm{i}p\int_0^t \frac{1}{a}\,\mathrm{d}t} + c_2(p)\mathrm{e}^{\mathrm{i}p\int_0^t \frac{1}{a}\,\mathrm{d}t} \right]
\end{aligned}$$

where the terms that go as $p^{-3/2}$ have been dropped since as $p \to \infty$ they go to zero faster than the terms that go as $p^{-1/2}$. To compare with the Minkowski case, we now set $a(t) = 1$ and get

---

[9] Compare this result to [60, page 158] and [94, equation (3.5)] but note that Linde omits the normalization condition on the $c_i(p)$ and that at this point I really can't say where he intends factors of H to be placed in his notation. The differential equation in the form he provides it in requires $a(t) = \mathrm{e}^{Ht}$, i.e. with no prefactor of H in it; for his solution to agree with this one, he must consider $\eta = -\mathrm{e}^{-Ht}$, i.e. to also have no prefactor of H in it; these two requirements, however, are incompatible with the requisite relationship between this two which is $\mathrm{d}\eta = \frac{1}{a}\,\mathrm{d}t$.

$$\psi_p(\eta) = \frac{1}{\sqrt{2p}}\left[c_1(p)\mathrm{e}^{-\mathrm{i}pt} + c_2(p)\mathrm{e}^{\mathrm{i}pt}\right]. \tag{C.7.15}$$

To have agreement with the Minkowski solution, we clearly need $c_1(p) \to 1$ and $c_2(p) \to 0$ as $p \to \infty$.[10]

C.7.1 Notes on Bessel Functions

For reference, the information summarized below can be found in [15, Section 3.3.1.3.4] and [3, Sections 11.3 and 11.4]. Bessel's equation of order n is

$$x^2\frac{\mathrm{d}^2}{\mathrm{d}x^2}y + x\frac{\mathrm{d}}{\mathrm{d}x}y + (k^2x^2 - n^2)y = 0 \tag{C.7.16}$$

for which the complete solution is

$$y(x) = \begin{cases} c_1\,\mathrm{J}_n(kx) + c_2\,\mathrm{J}_{-n}(kx) & \text{if } n \in \text{Integers}, \\ c_1\,\mathrm{J}_n(kx) + c_2\,\mathrm{Y}_n(kx) & \text{if } n \notin \text{Integers}. \end{cases} \tag{C.7.17}$$

J_n is the Bessel function of the first kind and Y_n is the Bessel function of the second kind also called the Weber function or Neumann function. J_{-n} is obtained by replacing n with $-n$ in the definition of J_n and if Y_n is defined as

$$\mathrm{Y}_n(x) = \lim_{\nu\to n}\frac{\cos(\nu\pi)\,\mathrm{J}_\nu(x) - \mathrm{J}_{-\nu}(x)}{\sin(\nu\pi)} \tag{C.7.18}$$

then $\mathrm{J}_{-n} \equiv \mathrm{Y}_n$ for integer n. One defines the Hankel functions of the first and second kinds as

$$\mathrm{H}_n^{(1)}(x) = \mathrm{J}_n(x) + \mathrm{i}\,\mathrm{Y}_n(x), \tag{C.7.19}$$

$$\mathrm{H}_n^{(2)}(x) = \mathrm{J}_n(x) - \mathrm{i}\,\mathrm{Y}_n(x). \tag{C.7.20}$$

Since the Hankel functions are linear combinations of J_n and J_{-n} or of J_n and Y_n respectively they are clearly solutions of Bessel's equation. Furthermore, since the two kinds of Hankel functions are themselves linearly independent, a linear combination of them can be used to form a complete solution of Bessel's equation

$$y(x) = c_1\,\mathrm{H}_n^{(1)}(x) + c_2\,\mathrm{H}_n^{(2)}(x). \tag{C.7.21}$$

---

[10] Compare this to the result that Linde provides on [60, page 158] but take note of several (further) quirks in Linde's result. He has the Hankel functions reversed relative to this document so what he calls c_1 and c_2, I call c_2 and c_1 respectively. See below for references regarding the definitions of the Hankel functions. In addition to that he has also made some sort of a sign error since even with his choice of Hankel functions, setting $c_1 = 0$ and $c_2 = -1$ as he suggests gives a result that disagrees with the Minkowski solution by an overall sign.

For real-valued arguments, the Hankel functions of the first and second kind are the complex conjugates of each other and for order 3/2 their explicit form is

$$\begin{aligned} \mathrm{H}^{(1)}_{3/2}(x) &= \mathrm{J}_{3/2}(x) + \mathrm{i}\,\mathrm{J}_{-3/2}(x) \\ &= \sqrt{\frac{2}{\pi x}}\left(\frac{1}{x}\sin x - \cos x\right) + \mathrm{i}\sqrt{\frac{2}{\pi x}}\left(\sin x + \frac{1}{x}\cos x\right) \\ &= \sqrt{\frac{2}{\pi x}}\left[\frac{\mathrm{i}}{x}\left(\cos x - \mathrm{i}\sin x\right) - \left(\cos x - \mathrm{i}\sin x\right)\right] \\ \left[\mathrm{H}^{(2)}_{3/2}(x)\right]^* = \mathrm{H}^{(1)}_{3/2}(x) &= -\sqrt{\frac{2}{\pi x}}\left(1 + \frac{1}{\mathrm{i}x}\right)\mathrm{e}^{-\mathrm{i}x}. \end{aligned} \tag{C.7.22}$$

The Bessel functions listed above, $\mathrm{J}_n(x)$, $\mathrm{Y}_n(x)$, etc., are often called the "cylinder" functions when n is an integer since these arise in the solution of the Laplace equation in cylindrical co-ordinates. Solving the Laplace equation in spherical co-ordinates gives rise to Bessel functions with half-odd-integer order, or "spherical" Bessel functions. For simplicity, a short-hand notation is introduced for this special case. The spherical Bessel functions are written as

$$\mathrm{j}_n(x) = \sqrt{\frac{\pi}{2x}}\ \mathrm{J}_{n+\frac{1}{2}}(x) \tag{C.7.23}$$

$$\mathrm{y}_n(x) = \sqrt{\frac{\pi}{2x}}\ \mathrm{Y}_{n+\frac{1}{2}}(x) \tag{C.7.24}$$

$$\mathrm{h}_n(x) = \sqrt{\frac{\pi}{2x}}\ \mathrm{H}_{n+\frac{1}{2}}(x). \tag{C.7.25}$$

C.8 The Derivation of (4.1.18)

The vacuum expectation value of φ^2 is

$$\begin{aligned} \left\langle\varphi^2\right\rangle &= \left\langle 0\left|\varphi^2\right|0\right\rangle \\ &= \frac{1}{(2\pi)^3}\left\langle 0\left|\int\left[a^\dagger_p\psi_p\mathrm{e}^{\mathrm{i}\vec{p}\cdot\vec{x}} + a_p\psi^*_p\mathrm{e}^{-\mathrm{i}\vec{p}\cdot\vec{x}}\right]\left[a^\dagger_{p'}\psi_{p'}\mathrm{e}^{\mathrm{i}\vec{p}'\cdot\vec{x}} + a_{p'}\psi^*_{p'}\mathrm{e}^{-\mathrm{i}\vec{p}'\cdot\vec{x}}\right]\mathrm{d}^3p\,\mathrm{d}^3p'\right|0\right\rangle \\ &= \frac{1}{(2\pi)^3}\int\left\langle 0\left|\left[a^\dagger_p\psi_p\mathrm{e}^{\mathrm{i}\vec{p}\cdot\vec{x}} + a_p\psi^*_p\mathrm{e}^{-\mathrm{i}\vec{p}\cdot\vec{x}}\right]\left[a^\dagger_{p'}\psi_{p'}\mathrm{e}^{\mathrm{i}\vec{p}'\cdot\vec{x}} + a_{p'}\psi^*_{p'}\mathrm{e}^{-\mathrm{i}\vec{p}'\cdot\vec{x}}\right]\right|0\right\rangle\mathrm{d}^3p\,\mathrm{d}^3p' \\ &= \frac{1}{(2\pi)^3}\int\left\langle 0\left|a_pa^\dagger_{p'}\right|0\right\rangle\psi_p\psi^*_{p'}\mathrm{e}^{-\mathrm{i}(\vec{p}-\vec{p}')\cdot\vec{x}}\,\mathrm{d}^3p\,\mathrm{d}^3p' \\ &= \frac{1}{(2\pi)^3}\int\left\langle 0\left|\delta\left(\vec{p}-\vec{p}'\right) + a^\dagger_{p'}a_p\right|0\right\rangle\psi_p\psi^*_{p'}\mathrm{e}^{-\mathrm{i}(\vec{p}-\vec{p}')\cdot\vec{x}}\,\mathrm{d}^3p\,\mathrm{d}^3p' \\ &= \frac{1}{(2\pi)^3}\int\delta\left(\vec{p}-\vec{p}'\right)\psi_p\psi^*_{p'}\mathrm{e}^{-\mathrm{i}(\vec{p}-\vec{p}')\cdot\vec{x}}\,\mathrm{d}^3p\,\mathrm{d}^3p' \\ \left\langle\varphi^2\right\rangle &= \frac{1}{(2\pi)^3}\int\left|\psi_p(t)\right|^2\,\mathrm{d}^3p \end{aligned} \tag{C.8.1}$$

where the commutation relation for the creation and annihilation operators,

$$a_{\vec{p}'} a^{\dagger}_{\vec{p}} - a^{\dagger}_{\vec{p}} a_{\vec{p}'} = \delta\left(\vec{p}' - \vec{p}\right), \tag{C.8.2}$$

and the properties

$$a_p \left|0\right\rangle = 0 \tag{C.8.3}$$

$$\left\langle 0\right| a^{\dagger}_p = 0 \tag{C.8.4}$$

$$\left\langle 0|0\right\rangle = 1 \tag{C.8.5}$$

have been used.[11]

C.9 Linear Cosmological Perturbations

C.9.1 Equations of Motion for the Newtonian Potential and Inflaton Perturbations

The ingredients for this derivation consist of the following: the unperturbed equations of motion for the Robertson-Walker scale factor which are given by (2.1.6a) and (2.1.6b); the equation of motion for the inflaton field given by (2.2.2); the stress-energy tensor for the inflaton field given by (2.2.3); and the metric for linear Newtonian perturbations in a flat FRW space-time give by (4.3.19). We will be considering the minimally-coupled case, so $\xi = 0$ in (2.2.2) and (2.2.3) but the calculations in this section do not depend on the precise form of $V(\phi)$ so long as it has no dependence on $g_{\mu\nu}$. For generality, $V(\phi)$ will remain unspecified apart from this restriction.

The Einstein Tensor

The first step we will take in the linearization of Einstein's field equation will be to obtain the Einstein tensor to linear order in Φ. This will be done using the procedure described in Appendix C.1. The metric is given by (4.3.19) and since it is diagonal we can use the simplified expressions for the Christoffel symbols in (C.1.5) to get

---

[11] See [96, equation (4.2.5)].

$$
\begin{array}{l|l|l|l}
\Gamma^t{}_{tt}\approx\dot\Phi & \Gamma^x{}_{tt}\approx\frac{1}{a^2}\Phi_{,x} & \Gamma^y{}_{tt}\approx\frac{1}{a^2}\Phi_{,y} & \Gamma^z{}_{tt}\approx\frac{1}{a^2}\Phi_{,z} \\
\Gamma^t{}_{tx}\approx\Phi_{,x} & \Gamma^x{}_{tx}\approx\frac{\dot a}{a}-\dot\Phi & \Gamma^y{}_{tx}=0 & \Gamma^z{}_{tx}=0 \\
\Gamma^t{}_{ty}\approx\Phi_{,y} & \Gamma^x{}_{ty}=0 & \Gamma^y{}_{ty}\approx\frac{\dot a}{a}-\dot\Phi & \Gamma^z{}_{ty}=0 \\
\Gamma^t{}_{tz}\approx\Phi_{,z} & \Gamma^x{}_{tz}=0 & \Gamma^y{}_{tz}=0 & \Gamma^z{}_{tz}\approx\frac{\dot a}{a}-\dot\Phi \\
\hline
\Gamma^t{}_{xt}\approx\Phi_{,x} & \Gamma^x{}_{xt}\approx\frac{\dot a}{a}-\dot\Phi & \Gamma^y{}_{xt}=0 & \Gamma^z{}_{xt}=0 \\
\Gamma^t{}_{xx}\approx(1-4\Phi)a^2\frac{\dot a}{a}-a^2\dot\Phi & \Gamma^x{}_{xx}\approx-\Phi_{,x} & \Gamma^y{}_{xx}\approx\Phi_{,y} & \Gamma^z{}_{xx}\approx\Phi_{,z} \\
\Gamma^t{}_{xy}=0 & \Gamma^x{}_{xy}\approx-\Phi_{,y} & \Gamma^y{}_{xy}\approx-\Phi_{,x} & \Gamma^z{}_{xy}=0 \\
\Gamma^t{}_{xz}=0 & \Gamma^x{}_{xz}\approx-\Phi_{,z} & \Gamma^y{}_{xz}=0 & \Gamma^z{}_{xz}\approx-\Phi_{,x} \\
\hline
\Gamma^t{}_{yt}\approx\Phi_{,y} & \Gamma^x{}_{yt}=0 & \Gamma^y{}_{yt}\approx\frac{\dot a}{a}-\dot\Phi & \Gamma^z{}_{yt}=0 \\
\Gamma^t{}_{yx}=0 & \Gamma^x{}_{yx}\approx-\Phi_{,y} & \Gamma^y{}_{yx}\approx-\Phi_{,x} & \Gamma^z{}_{yx}=0 \\
\Gamma^t{}_{yy}\approx(1-4\Phi)a^2\frac{\dot a}{a}-a^2\dot\Phi & \Gamma^x{}_{yy}\approx\Phi_{,x} & \Gamma^y{}_{yy}\approx-\Phi_{,y} & \Gamma^z{}_{yy}\approx\Phi_{,z} \\
\Gamma^t{}_{yz}=0 & \Gamma^x{}_{yz}=0 & \Gamma^y{}_{yz}\approx-\Phi_{,z} & \Gamma^z{}_{yz}\approx-\Phi_{,y} \\
\hline
\Gamma^t{}_{zt}\approx\Phi_{,z} & \Gamma^x{}_{zt}=0 & \Gamma^y{}_{zt}=0 & \Gamma^z{}_{zt}\approx\frac{\dot a}{a}-\dot\Phi \\
\Gamma^t{}_{zx}=0 & \Gamma^x{}_{zx}\approx-\Phi_{,z} & \Gamma^y{}_{zx}=0 & \Gamma^z{}_{zx}\approx-\Phi_{,x} \\
\Gamma^t{}_{zy}=0 & \Gamma^x{}_{zy}=0 & \Gamma^y{}_{zy}\approx-\Phi_{,z} & \Gamma^z{}_{zy}\approx-\Phi_{,y} \\
\Gamma^t{}_{zz}\approx(1-4\Phi)a^2\frac{\dot a}{a}-a^2\dot\Phi & \Gamma^x{}_{zz}\approx\Phi_{,x} & \Gamma^y{}_{zz}\approx\Phi_{,y} & \Gamma^z{}_{zz}\approx-\Phi_{,z}
\end{array}
\tag{C.9.1}
$$

where a dot indicates differentiation with respect to t. The first term in the Ricci tensor is

$$
\begin{aligned}
\Gamma^\alpha{}_{tt,\alpha} &= \ddot\Phi + \frac{1}{a^2}\nabla^2\Phi \\
\Gamma^\alpha{}_{ti,\alpha} = \Gamma^\alpha{}_{it,\alpha} &= 0 \\
\Gamma^\alpha{}_{ij,\alpha} &= \delta_{ij}a^2\left[-\ddot\Phi - 6\frac{\dot a}{a}\dot\Phi + (1-4\Phi)\left(\frac{\ddot a}{a} + \left(\frac{\dot a}{a}\right)^2\right)\right] + \delta_{ij}\nabla^2\Phi - 2\Phi_{,ij}.
\end{aligned}
$$

Contracting on the first and second indices of the Christoffel symbols gives

$$
\Gamma^\alpha{}_{t\alpha} = 3\frac{\dot a}{a} - 2\dot\Phi \qquad \Gamma^\alpha{}_{x\alpha} = -2\Phi_{,x} \qquad \Gamma^\alpha{}_{y\alpha} = -2\Phi_{,y} \qquad \Gamma^\alpha{}_{z\alpha} = -2\Phi_{,z} \tag{C.9.2}
$$

which allows easy calculation of the second

$$
\begin{aligned}
\Gamma^\alpha{}_{t\alpha,t} &= 3\left(\frac{\ddot a}{a} - \left(\frac{\dot a}{a}\right)^2\right) - 2\ddot\Phi \\
\Gamma^\alpha{}_{t\alpha,i} = \Gamma^\alpha{}_{i\alpha,t} &= -2\dot\Phi_{,i} \\
\Gamma^\alpha{}_{i\alpha,j} &= -2\Phi_{,ij}
\end{aligned}
$$

and the third terms of the Ricci tensor

$$
\Gamma^\alpha{}_{\beta\alpha}\Gamma^\beta{}_{tt} = 3\frac{\dot a}{a}\dot\Phi
$$

$$\begin{aligned}
{\Gamma^\alpha}_{\beta\alpha}{\Gamma^\beta}_{ti} = {\Gamma^\alpha}_{\beta\alpha}{\Gamma^\beta}_{it} &= \frac{\dot{a}}{a}\Phi_{,i} \\
{\Gamma^\alpha}_{\beta\alpha}{\Gamma^\beta}_{ij} &= \delta_{ij}a^2\left[-5\frac{\dot{a}}{a}\dot{\Phi} + 3(1-4\Phi)\left(\frac{\dot{a}}{a}\right)^2\right].
\end{aligned}$$

The fourth term is

$$\begin{aligned}
{\Gamma^\alpha}_{t\beta}{\Gamma^\beta}_{t\alpha} &= 3\left(\frac{\dot{a}}{a}\right)^2 - 6\frac{\dot{a}}{a}\dot{\Phi} \\
{\Gamma^\alpha}_{t\beta}{\Gamma^\beta}_{i\alpha} = {\Gamma^\alpha}_{i\beta}{\Gamma^\beta}_{t\alpha} &= -\frac{\dot{a}}{a}\Phi_{,i} \\
{\Gamma^\alpha}_{i\beta}{\Gamma^\beta}_{j\alpha} &= \delta_{ij}a^2\left[-4\frac{\dot{a}}{a}\dot{\Phi} + 2(1-4\Phi)\left(\frac{\dot{a}}{a}\right)^2\right].
\end{aligned}$$

Therefore, to linear order in Φ, the Ricci tensor is given by

$$\begin{aligned}
R_{tt} &= 3\ddot{\Phi} + 9\frac{\dot{a}}{a}\dot{\Phi} + \frac{1}{a^2}\nabla^2\Phi - 3\frac{\ddot{a}}{a} \\
R_{ti} = R_{it} &= 2\dot{\Phi}_{,i} + 2\frac{\dot{a}}{a}\Phi_{,i} \\
R_{ij} &= \delta_{ij}a^2\left[-\ddot{\Phi} - 7\frac{\dot{a}}{a}\dot{\Phi} + (1-4\Phi)\left(\frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2\right)\right] + \delta_{ij}\nabla^2\Phi.
\end{aligned}$$

The Ricci scalar is, therefore,

$$R = -6\ddot{\Phi} - 30\frac{\dot{a}}{a}\dot{\Phi} + 2\frac{1}{a^2}\nabla^2\Phi + 6(1-2\Phi)\left(\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2\right).$$

The quantity $-\frac{1}{2}g_{\mu\nu}R$ is

$$\begin{aligned}
-\frac{1}{2}g_{tt}R &= -3\ddot{\Phi} - 15\frac{\dot{a}}{a}\dot{\Phi} + \frac{1}{a^2}\nabla^2\Phi + 3\left(\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2\right) \\
-\frac{1}{2}g_{ti}R = -\frac{1}{2}g_{it}R &= 0 \\
-\frac{1}{2}g_{ij}R &= \delta_{ij}a^2\left[3\ddot{\Phi} + 15\frac{\dot{a}}{a}\dot{\Phi} - 3(1-4\Phi)\left(\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2\right)\right] - \delta_{ij}\nabla^2\Phi.
\end{aligned}$$

This now gives us the Einstein tensor to linear order in Φ and its derivatives

$$G_{tt} = 3\left(\frac{\dot{a}}{a}\right)^2 - 6\frac{\dot{a}}{a}\dot{\Phi} + 2\frac{1}{a^2}\nabla^2\Phi \tag{C.9.3a}$$

$$G_{ti} = G_{it} = 2\dot{\Phi}_{,i} + 2\frac{\dot{a}}{a}\Phi_{,i} \tag{C.9.3b}$$

$$G_{ij} = \delta_{ij} a^2 \left[2\ddot{\Phi} + 8\frac{\dot{a}}{a}\dot{\Phi} - (1-4\Phi)\left(2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2\right)\right] \tag{C.9.3c}$$

or, raising one index,[12]

$$G^t{}_t = -3H^2 + 6H^2\Phi + 6H\dot{\Phi} - 2\frac{1}{a^2}\nabla^2\Phi \tag{C.9.4a}$$

$$G^t{}_i = G^i{}_t = -2\dot{\Phi}_{,i} - 2H\Phi_{,i} \tag{C.9.4b}$$

$$G^i{}_j = \delta^i{}_j \left[2\ddot{\Phi} + 8H\dot{\Phi} - (1-2\Phi)\left(2\dot{H} + 3H^2\right)\right]. \tag{C.9.4c}$$

The Inflaton Stress-Energy Tensor

The second step in linearizing Einstein's equation is to obtain the stress-energy tensor for the scalar field to linear orders in the field perturbation $\delta\phi$ and the metric perturbation Φ. The stress-energy tensor is given by (2.2.3) and the first step will be to write it explicitly to linear order in Φ alone which is

$$\begin{aligned}
T_{\mu\nu} &= \phi_{;\mu}\phi_{;\nu} - \frac{1}{2}g^{\rho\sigma}\phi_{;\rho}\phi_{;\sigma}g_{\mu\nu} - g_{\mu\nu}V(\phi) \\
&= \phi_{,\mu}\phi_{,\nu} - \frac{1}{2}g_{\mu\nu}\left[-\frac{1}{1+2\Phi}\dot{\phi}^2 + \frac{1}{(1-2\Phi)a^2}(\nabla\phi)^2\right] - g_{\mu\nu}V(\phi) \\
&= \phi_{,\mu}\phi_{,\nu} + \frac{1}{2}g_{\mu\nu}\left[(1-2\Phi)\dot{\phi}^2 - (1+2\Phi)\frac{1}{a^2}(\nabla\phi)^2\right] - g_{\mu\nu}V(\phi)
\end{aligned}$$

so

$$\begin{aligned}
T_{tt} &= \frac{1}{2}\dot{\phi}^2 + \frac{1}{2}(1+4\Phi)\frac{1}{a^2}(\nabla\phi)^2 + (1+2\Phi)V(\phi) \\
T_{ti} = T_{it} &= \dot{\phi}\phi_{,i} \\
T_{ij} &= \phi_{,i}\phi_{,j} + \delta_{ij}a^2\left[(1-4\Phi)\frac{1}{2}\dot{\phi}^2 - \frac{1}{2}\frac{1}{a^2}(\nabla\phi)^2 - (1-2\Phi)V(\phi)\right].
\end{aligned}$$

Now we make the substitution $\phi \to \phi + \delta\phi$ and only keep terms up to and including those linear in the perturbations (i.e. taking terms like $\Phi\delta\phi$ to be second order) which gives

$$\begin{aligned}
T_{tt} &= \frac{1}{2}\dot{\phi}^2 + \dot{\phi}\delta\dot{\phi} + \frac{1}{2}(1+4\Phi)\frac{1}{a^2}(\nabla\phi)^2 + \frac{1}{a^2}\nabla\phi\cdot\nabla\delta\phi + (1+2\Phi)V(\phi) + \frac{\mathrm{d}V}{\mathrm{d}\phi}\delta\phi \\
T_{ti} = T_{it} &= \dot{\phi}\phi_{,i} + \delta\dot{\phi}\phi_{,i} + \dot{\phi}\delta\phi_{,i}
\end{aligned}$$

---

[12] This result agrees with the result obtained by Mukhanov et al. in [69, equation (4.11)]. When comparing results, it is helpful to use the identities in (C.9.16).

$$\begin{aligned}T_{ij} = \phi_{,i}\phi_{,j} + \delta\phi_{,i}\phi_{,j} + \phi_{,i}\delta\phi_{,j} + \\ \delta_{ij}a^2\left[\frac{1}{2}(1-4\Phi)\dot{\phi}^2 + \dot{\phi}\delta\dot{\phi} - \frac{1}{2}\frac{1}{a^2}\left[(\nabla\phi)^2 + 2\nabla\phi\cdot\nabla\delta\phi\right] - \right. \\ \left. (1-2\Phi)V(\phi) - \frac{\mathrm{d}V}{\mathrm{d}\phi}\delta\phi\right].\end{aligned}$$

Now, recall that ϕ is the background, unperturbed, field. Since we have assumed this solution to be homogeneous the spatial derivatives of ϕ are 0 so

$$\begin{aligned}T_{tt} &= \frac{1}{2}\dot{\phi}^2 + \dot{\phi}\delta\dot{\phi} + (1+2\Phi)V(\phi) + \frac{\mathrm{d}V}{\mathrm{d}\phi}\delta\phi \\ T_{ti} = T_{it} &= \dot{\phi}\delta\phi_{,i} \\ T_{ij} &= \delta_{ij}a^2\left[\frac{1}{2}(1-4\Phi)\dot{\phi}^2 + \dot{\phi}\delta\dot{\phi} - (1-2\Phi)V(\phi) - \frac{\mathrm{d}V}{\mathrm{d}\phi}\delta\phi\right].\end{aligned}$$

For reasons that will become clear later, we'll rewrite this as

$$\begin{aligned}T_{tt} &= \frac{3}{2}\left(\frac{1}{3}\dot{\phi}^2 + \frac{2}{3}V(\phi)\right) + 2\Phi V(\phi) + \dot{\phi}\delta\dot{\phi} + \frac{\mathrm{d}V}{\mathrm{d}\phi}\delta\phi \\ T_{ti} = T_{it} &= \dot{\phi}\delta\phi_{,i} \\ T_{ij} &= \delta_{ij}a^2\left[-\frac{1}{2}(1-4\Phi)\left(-\dot{\phi}^2 + 2V(\phi)\right) - 2\Phi V(\phi) + \dot{\phi}\delta\dot{\phi} - \frac{\mathrm{d}V}{\mathrm{d}\phi}\delta\phi\right].\end{aligned}$$

Einstein's Equation

Einstein's equation is $G_{\mu\nu} = 8\pi T_{\mu\nu}$ from which we get the following three equations to linear order in Φ and $\delta\phi$

$$\begin{aligned}3\frac{\dot{a}}{a}\dot{\Phi} - \frac{1}{a^2}\nabla^2\Phi - \frac{3}{2}\left(\frac{\dot{a}}{a}\right)^2 &= -\frac{3}{2}\left(\frac{4\pi}{3}\dot{\phi}^2 + \frac{8\pi}{3}V(\phi)\right) - 8\pi\Phi V(\phi) - 4\pi\dot{\phi}\delta\dot{\phi} - 4\pi\frac{\mathrm{d}V}{\mathrm{d}\phi}\delta\phi \\ \dot{\Phi}_{,i} + \frac{\dot{a}}{a}\Phi_{,i} &= 4\pi\dot{\phi}\delta\phi_{,i} \\ \ddot{\Phi} + 4\frac{\dot{a}}{a}\dot{\Phi} - \frac{1}{2}(1-4\Phi)\left(2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2\right) & \\ = -\frac{1}{2}(1-4\Phi)\left(-4\pi\dot{\phi}^2 + 8\pi V(\phi)\right) &- 8\pi\Phi V(\phi) + 4\pi\dot{\phi}\delta\dot{\phi} - 4\pi\frac{\mathrm{d}V}{\mathrm{d}\phi}\delta\phi.\end{aligned}$$

The unperturbed equations of motion for $a(t)$ are

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi}{3}\dot{\phi}^2 + \frac{8\pi}{3}V(\phi) \tag{C.9.5a}$$

$$2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 = -4\pi\dot{\phi}^2 + 8\pi V(\phi) \tag{C.9.5b}$$

which can be used to show that

$$\frac{\ddot{a}}{a} - \left(\frac{\dot{a}}{a}\right)^2 = \left(\frac{\dot{a}}{a}\right)^{\cdot} = -4\pi\dot{\phi}^2 \tag{C.9.6a}$$

$$\frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2 = 8\pi V(\phi). \tag{C.9.6b}$$

Imposing the requirement that $a(t)$ satisfy its unperturbed equations of motion reduces the perturbed equations to[13]

$$3\frac{\dot{a}}{a}\dot{\Phi} - \frac{1}{a^2}\nabla^2\Phi = \left(4\pi\dot{\phi}^2 - 3\left(\frac{\dot{a}}{a}\right)^2\right)\Phi - 4\pi\dot{\phi}\delta\dot{\phi} - 4\pi\frac{\mathrm{d}V}{\mathrm{d}\phi}\delta\phi \tag{C.9.7a}$$

$$\dot{\Phi}_{,i} + \frac{\dot{a}}{a}\Phi_{,i} = 4\pi\dot{\phi}\delta\phi_{,i} \tag{C.9.7b}$$

$$\ddot{\Phi} + 4\frac{\dot{a}}{a}\dot{\Phi} = \left(-4\pi\dot{\phi}^2 - 2\frac{\ddot{a}}{a} - \left(\frac{\dot{a}}{a}\right)^2\right)\Phi + 4\pi\dot{\phi}\delta\dot{\phi} - 4\pi\frac{\mathrm{d}V}{\mathrm{d}\phi}\delta\phi. \tag{C.9.7c}$$

Expanding Φ and $\delta\phi$ into plane wave modes in (C.9.7b) gives

$$ik\dot{\Phi}_k + \frac{\dot{a}}{a}ik\Phi_k = 4\pi\dot{\phi}ik\delta\phi_k \tag{C.9.8}$$

but noting that $\Phi_{k=0}$ and $\delta\phi_{k=0}$ are both $= 0$ since those terms are accounted for by the homogeneous background allows us to divide both sides by ik to get

$$\dot{\Phi}_k + \frac{\dot{a}}{a}\Phi_k = 4\pi\dot{\phi}\delta\phi_k. \tag{C.9.9}$$

Inverting the mode expansion then gives simply

$$\dot{\Phi} + \frac{\dot{a}}{a}\Phi = 4\pi\dot{\phi}\delta\phi. \tag{C.9.10}$$

Adding (C.9.7a) to (C.9.7c) in order to eliminate both $\dot{\phi}$ and $\delta\dot{\phi}$ leaves

$$\ddot{\Phi} + 7\frac{\dot{a}}{a}\dot{\Phi} - \frac{1}{a^2}\nabla^2\Phi + 2\left(\frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2\right)\Phi = -8\pi\frac{\mathrm{d}V}{\mathrm{d}\phi}\delta\phi \tag{C.9.11}$$

---

[13] Equations (C.9.7a) and (C.9.7b) agree with [68, equations (8), and (9)] respectively, and also with [69, equations (6.40) and (6.41)] respectively when [69, equation (6.43)] is used. The third, (C.9.7c), agrees with [69, equation (6.42)] but differs from [68, equation (10)] by $-2H^2\Phi$. The error in Mukhanov's paper is probably a result of the hand-drawn parentheses in the term $2\left(\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2\right)\Phi$ which, if drawn as $\left(2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2\right)\Phi$, fix the problem.

and imposing the condition that ϕ obey its unperturbed equation of motion which is given by (2.3.11) with the spatial derivatives set to zero i.e.

$$\ddot{\phi} + 3\frac{\dot{a}}{a}\dot{\phi} = -\frac{\mathrm{d}V}{\mathrm{d}\phi} \tag{C.9.12}$$

gives us

$$\ddot{\Phi} + 7\frac{\dot{a}}{a}\dot{\Phi} - \frac{1}{a^2}\nabla^2\Phi + 2\left(\frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2\right)\Phi = 8\pi\dot{\phi}\delta\phi\left(\frac{\ddot{\phi}}{\dot{\phi}} + 3\frac{\dot{a}}{a}\right). \tag{C.9.13}$$

Using (C.9.10) for $4\pi\delta\phi$ in (C.9.13) leaves

$$\ddot{\Phi} + 7\frac{\dot{a}}{a}\dot{\Phi} - \frac{1}{a^2}\nabla^2\Phi + 2\left(\frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2\right)\Phi = 2\left(\dot{\Phi} + \frac{\dot{a}}{a}\Phi\right)\left(\frac{\ddot{\phi}}{\dot{\phi}} + 3\frac{\dot{a}}{a}\right). \tag{C.9.14}$$

If, instead of adding (C.9.7a) and (C.9.7c) together as was done above to obtain (C.9.14), one subtracts (C.9.7c) from (C.9.7a) one will in fact simply obtain (C.9.14) again but by a different route. In other words, the four equations in (C.9.7) and (C.9.12) are not all independent. Rewriting (C.9.14), Einstein's field equation leaves us with only the following two independent, non-trivial, equations[14]

$$\ddot{\Phi} + \left(\frac{\dot{a}}{a} - 2\frac{\ddot{\phi}}{\dot{\phi}}\right)\dot{\Phi} - \frac{1}{a^2}\nabla^2\Phi + 2\left(\frac{\ddot{a}}{a} - \left(\frac{\dot{a}}{a}\right)^2 - \frac{\dot{a}}{a}\frac{\ddot{\phi}}{\dot{\phi}}\right)\Phi = 0 \tag{C.9.15a}$$

$$\frac{1}{a}\dot{(a\Phi)}_{,i} = 4\pi(\dot{\phi}\delta\phi)_{,i}. \tag{C.9.15b}$$

Conformal Time

By switching to conformal time, η, defined by (2.1.5), it is possible to rewrite (C.9.15a) in a more compact form. This can done as follows although one almost needs to know the answer in order to find it. We will need the following relationships where $\mathcal{H} = \frac{a'}{a}$ and $'$ denotes differentiation with respect to conformal time.

$$\frac{\dot{a}}{a} = \frac{1}{a}\mathcal{H} \qquad \frac{\ddot{a}}{a} = \frac{1}{a^2}\mathcal{H}' \qquad \dot{X} = \frac{1}{a}X' \qquad \ddot{X} = \frac{1}{a^2}X'' - \frac{1}{a^2}\mathcal{H}X' \tag{C.9.16}$$

---

[14] These equations agree with [69, equations (6.48) and (6.41)] respectively and also with [68, equations (11) and (9)] respectively, confirming the assertion made in footnote 13 above that Mukhanov has a typographic error in [68, equation (10)]. These equations do not, however, agree with [60, equations (7.5.10) and (7.5.11)] where Linde has incorrectly transcribed Mukhanov's results verbatim: throughout his book Linde uses Greek indices to mean all co-ordinates while in [68] Mukhanov uses Greek indices to mean only spatial co-ordinates.

These are easily verified. Using (C.9.16), we can rewrite the background equation, (C.9.6a), as[15]

$$\mathcal{H}^2 - \mathcal{H}' = 4\pi\phi'^2. \tag{C.9.17}$$

Applying (C.9.16) to the equation of motion in (C.9.15a) gives[16]

$$\Phi'' + 2\left(\mathcal{H} - \frac{\phi''}{\phi'}\right)\Phi' - \nabla^2\Phi + 2\left(\mathcal{H}' - \mathcal{H}\frac{\phi''}{\phi'}\right)\Phi = 0. \tag{C.9.18}$$

Now, introducing

$$u = \frac{a}{\phi'}\Phi, \tag{C.9.19a}$$

we find that

$$\nabla^2 u = \frac{a}{\phi'}\nabla^2\Phi \tag{C.9.19b}$$

and

$$\begin{aligned} u'' &= \left\{\frac{a}{\phi'}\left[\Phi' + \left(\mathcal{H} - \frac{\phi''}{\phi'}\right)\Phi\right]\right\}' \\ &= \frac{a}{\phi'}\left\{\Phi'' + 2\left(\mathcal{H} - \frac{\phi''}{\phi'}\right)\Phi' + \left[\mathcal{H}^2 + \mathcal{H}' - 2\mathcal{H}\frac{\phi''}{\phi'} + 2\left(\frac{\phi''}{\phi'}\right)^2 - \frac{\phi'''}{\phi'}\right]\Phi\right\}. \end{aligned} \tag{C.9.19c}$$

From (C.9.19a), (C.9.19b) and (C.9.19c), we see that the equation of motion in (C.9.18) can be written as

$$u'' - \nabla^2 u - \left[\mathcal{H}^2 - \mathcal{H}' + 2\left(\frac{\phi''}{\phi'}\right)^2 - \frac{\phi'''}{\phi'}\right]u = 0. \tag{C.9.20}$$

Now,

$$\begin{aligned} \left(\frac{\mathcal{H}}{a\phi'}\right)'' &= \left(\frac{1}{a\phi'}\mathcal{H}' - \frac{1}{a\phi'}\mathcal{H}^2 - \frac{\phi''}{a\phi'^2}\mathcal{H}\right)' \\ &= \left(-4\pi\frac{\phi'}{a} - \frac{\phi''}{a\phi'^2}\mathcal{H}\right)' \\ &= 4\pi\left(\frac{\phi'}{a}\mathcal{H} - \frac{\phi''}{a}\right) - \frac{\phi'''}{a\phi'^2}\mathcal{H} - \frac{\phi''}{a\phi'^2}\mathcal{H}' + \frac{\phi''}{a\phi'^2}\mathcal{H}^2 + 2\frac{\phi''^2}{a\phi'^3}\mathcal{H} \\ &= (\mathcal{H}^2 - \mathcal{H}')\left(\frac{1}{a\phi'}\mathcal{H} - \frac{\phi''}{a\phi'^2}\right) - \frac{\phi'''}{a\phi'^2}\mathcal{H} - \frac{\phi''}{a\phi'^2}\mathcal{H}' + \frac{\phi''}{a\phi'^2}\mathcal{H}^2 + 2\frac{\phi''^2}{a\phi'^3}\mathcal{H} \\ \left(\frac{\mathcal{H}}{a\phi'}\right)'' &= \left(\frac{\mathcal{H}}{a\phi'}\right)\left[\mathcal{H}^2 - \mathcal{H}' - \frac{\phi'''}{\phi'} + 2\left(\frac{\phi''}{\phi'}\right)^2\right], \end{aligned} \tag{C.9.21}$$

---

[15] This is identical to [69, equation (6.43)].

[16] This is identical to [69, equation (6.48)].

where (C.9.17) has been used twice. From this, the equation of motion in (C.9.20) can be written as

$$u'' - \nabla^2 u - \frac{\left[a'/(a^2\phi')\right]''}{\left[a'/(a^2\phi')\right]} u = 0 \tag{C.9.22}$$

where $u = \frac{a}{\phi'}\Phi$ and a prime indicates differentiation with respect to conformal time, η.[17] While this equation is third-order in the time derivatives of ϕ it should be remembered that ϕ is the background and so it and its derivatives appear purely parametrically in this equation which is still only a linear second-order differential equation for the metric perturbation — the quantity whose dynamics are being expressed.

Euler-Lagrange Equation

The final thing we need to do is linearize the equation of motion for ϕ in both Φ and $\delta\phi$. The equation of motion is (2.2.2) which can be written as

$$\Box\phi = \frac{\mathrm{d}V}{\mathrm{d}\phi}. \tag{C.9.23}$$

Starting from (C.5.4) we obtain the following for the d'Alembertian with respect to our perturbed line element.

$$\begin{aligned}
\Box\phi &= g^{00}\left(\phi_{,00} - {\Gamma^\alpha}_{00}\phi_{,\alpha}\right) - g^{ii}{\Gamma^0}_{ii}\phi_{,0} + g^{ij}\left(\phi_{,ij} - {\Gamma^\alpha}_{ij}\phi_{,\alpha}\right) \\
&= g^{00}\left(\phi_{,00} - {\Gamma^\alpha}_{00}\phi_{,\alpha}\right) - g^{ii}{\Gamma^0}_{ii}\phi_{,0} + \frac{1}{(1-2\Phi)a^2}\nabla^2\phi \\
&= -\frac{1}{1+2\Phi}\left(\ddot{\phi} - \dot{\Phi}\dot{\phi} - \frac{1}{a^2}\nabla\Phi\cdot\nabla\phi\right) - \\
&\qquad\qquad \frac{3}{(1-2\Phi)a^2}\left[-a^2\dot{\Phi} + (1-4\Phi)a^2\frac{\dot{a}}{a}\right]\dot{\phi} + \frac{1}{(1-2\Phi)a^2}\nabla^2\phi
\end{aligned}$$

$$\Box\phi = -(1-2\Phi)\left(\ddot{\phi} + 3\frac{\dot{a}}{a}\dot{\phi}\right) + 4\dot{\Phi}\dot{\phi} + \frac{1}{a^2}\nabla\Phi\cdot\nabla\phi + (1+2\Phi)\frac{1}{a^2}\nabla^2\phi \tag{C.9.24}$$

and so (C.9.23) to linear order in Φ reads as

$$(1-2\Phi)\left(\ddot{\phi} + 3\frac{\dot{a}}{a}\dot{\phi}\right) - 4\dot{\Phi}\dot{\phi} - \frac{1}{a^2}\nabla\Phi\cdot\nabla\phi - (1+2\Phi)\frac{1}{a^2}\nabla^2\phi = -\frac{\mathrm{d}V}{\mathrm{d}\phi}.$$

Making the substitution $\phi \to \phi + \delta\phi$ and keeping terms up to and including those of linear order in the perturbations (i.e. taking terms like $\Phi\delta\phi$ to be second order) gives

---

[17] This result agrees with [68, equation (12)] and [69, equation (6.51)]. It does not agree with [60, equation (7.5.14)] where Linde has added an extra derivative in the numerator of the u term. Also, in the case of [68], Mukhanov has been neither careful nor consistent with his order of operations: in the expression he writes as $a'/a^2\varphi_0'$ the φ_0' is meant to be in the denominator while immediately above that where he has written $a/\varphi'\Phi$ the Φ is meant to be in the numerator. Only when this context-sensitive order of operations is used does his result agree with the one here.

$$(1-2\Phi)\left(\ddot{\phi}+3\frac{\dot{a}}{a}\dot{\phi}\right)+\delta\ddot{\phi}+3\frac{\dot{a}}{a}\delta\dot{\phi}-4\dot{\Phi}\dot{\phi}-\frac{1}{a^2}\nabla\Phi\cdot\nabla\phi-$$
$$(1+2\Phi)\frac{1}{a^2}\nabla^2\phi-\frac{1}{a^2}\Delta\delta\phi+\frac{\mathrm{d}V}{\mathrm{d}\phi}+\frac{\mathrm{d}^2V}{\mathrm{d}\phi^2}\delta\phi=0.$$

Imposing the requirement that ϕ satisfy the unperturbed equation of motion, (C.9.12), and setting any remaining spatial derivatives of ϕ to zero leaves a final result of[18]

$$\delta\ddot{\phi}+3\frac{\dot{a}}{a}\delta\dot{\phi}-\frac{1}{a^2}\nabla^2\delta\phi+\frac{\mathrm{d}^2V}{\mathrm{d}\phi^2}\delta\phi+2\frac{\mathrm{d}V}{\mathrm{d}\phi}\Phi-4\dot{\Phi}\dot{\phi}=0. \tag{C.9.25}$$

Equations (C.9.15a), (C.9.15b), and (C.9.25) constitute the linearized equations of motion for the perturbations.

C.9.2 Equations of Motion for the Newtonian Potential and Dust Perturbations

Einstein's Equation

Here we will briefly examine the behaviour of perturbations when the source matter consists of linearly perturbed dust rather than a linearly perturbed inflaton field. In this case, the left-hand side of Einstein's equation is the same as before so the Einstein tensor in (C.9.3) can be reused. The stress-energy tensor for a perfect fluid is given in (2.1.2) and using the perturbed metric tensor from (4.3.19) it becomes

$$T_{tt}=(\rho+p){u_t}^2-(1+2\Phi)\,p \tag{C.9.26a}$$
$$T_{ti}=T_{it}=(\rho+p)u_tu_i \tag{C.9.26b}$$
$$T_{ij}=(\rho+p)u_iu_j+(1-2\Phi)\,a^2p\delta_{ij}. \tag{C.9.26c}$$

The background cosmological fluid is co-moving. Introducing small perturbations to the components of its four-velocity, we can write[19]

$$u^t=1-\Phi \tag{C.9.27a}$$
$$u^i=\delta u_i \tag{C.9.27b}$$

where $|\delta u_i|\ll 1$. The time component above is obtained by imposing the usual normalization for a four-velocity,[20]

$$g_{\mu\nu}u^\mu u^\nu=-1, \tag{C.9.28}$$

which is carried out using the perturbed metric and four-velocity but only computed to first order in the perturbations. A particle observed to have a proper three velocity of

---

[18] Compare this to [60, equation (7.5.12)].

[19] Note that the spatial components alone do not constitute a tensor so the concepts of "up" and "down" indices are irrelevant when writing them.

[20] See [81, equation (2.28)].

v_i in the cosmological rest frame has, for the spatial components of its four-velocity,[21]

$$a\delta u_i = v_i/\sqrt{1-v^2} \approx v_i. \tag{C.9.29}$$

The covariant components of the four velocity are

δ

$$u_t = -(1+\Phi) \tag{C.9.30a}$$
$$u_i = a^2\delta u_i. \tag{C.9.30b}$$

Using the perturbed four-velocity from (C.9.30) in the stress-energy tensor, (C.9.26), and writing the result to first order gives

$$\begin{aligned} T_{tt} &= (1+2\Phi)\,\rho \\ T_{ti} = T_{it} &= -(\rho+p)a^2\delta u_i \\ T_{ij} &= (1-2\Phi)\,a^2 p\delta_{ij}. \end{aligned}$$

Finally, introducing a perturbed energy density, $\rho \to \rho_{\mathrm{b}} + \delta\rho$, and setting the pressure to zero (we are considering dust) gives, to linear order in the perturbations,

$$T_{tt} = (1+2\Phi)\,\rho_{\mathrm{b}} + \delta\rho \tag{C.9.31a}$$
$$T_{ti} = T_{it} = -\rho_{\mathrm{b}} a^2\delta u_i \tag{C.9.31b}$$
$$T_{ij} = 0. \tag{C.9.31c}$$

The three-vector describing the perturbation to the spatial part of the four-velocity cannot be chosen arbitrarily. It must be remembered that our analysis is being confined to scalar perturbations and this puts restrictions on the form of the perturbations to the stress-energy tensor. In particular, the spatial components of the four velocity perturbation must be expressible as the gradient of a scalar,[22]

$$\delta u_i = U_{,i}. \tag{C.9.32}$$

This condition will not be imposed at this time, instead we will see it come out as a consequence of Einstein's equation. Equating the Einstein tensor from (C.9.3) to 8π times the stress-energy tensor from (C.9.31) gives Einstein's equation linearized with respect to the perturbations,

$$3\left(\frac{\dot a}{a}\right)^2 - 6\frac{\dot a}{a}\dot\Phi + 2\frac{1}{a^2}\nabla^2\Phi = 8\pi\,(1+2\Phi)\,\rho_{\mathrm{b}} + 8\pi\delta\rho \tag{C.9.33a}$$
$$2\dot\Phi_{,i} + 2\frac{\dot a}{a}\Phi_{,i} = -8\pi\rho_{\mathrm{b}} a^2\delta u_i \tag{C.9.33b}$$

---

[21] See [81, Section 2.3].

[22] Compare to [69, equation (5.1)].

$$\ddot{\Phi} + 4\frac{\dot{a}}{a}\dot{\Phi} - \frac{1}{2}(1-4\Phi)\left(2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2\right) = 0 \tag{C.9.33c}$$

The unperturbed equations of motion are given by (2.1.6) with $k = 0$ and $p = 0$,

$$2\left(\frac{\ddot{a}}{a}\right) + \left(\frac{\dot{a}}{a}\right)^2 = 0, \tag{C.9.34a}$$

$$\rho_b = \frac{3}{8\pi}\left(\frac{\dot{a}}{a}\right)^2. \tag{C.9.34b}$$

From (C.9.34a) it can be shown that

$$\left(\frac{\dot{a}}{a}\right)^{\cdot} = -\frac{3}{2}\left(\frac{\dot{a}}{a}\right)^2. \tag{C.9.35}$$

Imposing the requirement that the unperturbed equations in (C.9.34) be satisfied turns (C.9.33) into[23]

$$\frac{1}{a^2}\nabla^2\Phi - 3\frac{\dot{a}}{a}\dot{\Phi} - 3\left(\frac{\dot{a}}{a}\right)^2\Phi = 4\pi\delta\rho \tag{C.9.36a}$$

$$2\dot{\Phi}_{,i} + 2\frac{\dot{a}}{a}\Phi_{,i} = -3\left(\frac{\dot{a}}{a}\right)^2 a^2\delta u_i \tag{C.9.36b}$$

$$\ddot{\Phi} + 4\frac{\dot{a}}{a}\dot{\Phi} = 0. \tag{C.9.36c}$$

Metric Perturbations

In a critical-density universe, (2.1.13) tells us that

$$\frac{\dot{a}}{a} = \frac{2}{3(1+w)}t^{-1},$$

and in the absence of pressure ($w = 0$) this reduces to $\frac{\dot{a}}{a} = \frac{2}{3}t^{-1}$, so (C.9.36c) can be written as

$$\ddot{\Phi} + \frac{8}{3t}\dot{\Phi} = 0.$$

It can be verified that the general solution to this is[24]

$$\Phi(\vec{x}, t) = \Phi_1(\vec{x}) + \Phi_2(\vec{x})t^{-\frac{5}{3}}, \tag{C.9.37}$$

---

[23] These agree with [69, equations (5.17) – (5.19)] when their p_0 and δp are both $= 0$. Note that in their equations, $2\mathcal{H}' + \mathcal{H}^2$ forms the left-hand side of the spatial part of the unperturbed Einstein equation in a $\mathcal{K} = 0$ universe (see their equation (4.2)) and so is equal to 0 in the absence of pressure.

[24] Agrees with [69, equation (5.32)].

where Φ_1 and Φ_2 are arbitrary functions of position. The conclusion from this is that in a pressureless, flat, universe, the metric perturbation quickly (within several Planck times) becomes approximately independent of time.

Velocity Perturbations

From the components of the Einstein equation in (C.9.36) we can obtain the relationship between the metric and four velocity perturbations. Differentiating (C.9.36b) with respect to time gives

$$2\ddot{\Phi}_{,i} + 2\left(\frac{\dot{a}}{a}\right)^{\cdot} \Phi_{,i} + 2\left(\frac{\dot{a}}{a}\right)\dot{\Phi}_{,i} = -6\left(\frac{\dot{a}}{a}\right)\left(\frac{\dot{a}}{a}\right)^{\cdot} a^2\delta u_i - 3\left(\frac{\dot{a}}{a}\right)^2 (a^2\dot{\delta u_i})$$

and using (C.9.35) this becomes

$$2\ddot{\Phi}_{,i} - 3\left(\frac{\dot{a}}{a}\right)^2 \Phi_{,i} + 2\left(\frac{\dot{a}}{a}\right)\dot{\Phi}_{,i} = 9\left(\frac{\dot{a}}{a}\right)^3 a^2\delta u_i - 3\left(\frac{\dot{a}}{a}\right)^2 (a^2\dot{\delta u_i}). \qquad \text{(C.9.38)}$$

The result of (C.9.38) $-\, 2 \times$ (C.9.36c)$_{,i} + 3(\frac{\dot{a}}{a}) \times$ (C.9.36b) is

$$3\left(\frac{\dot{a}}{a}\right)^2 \Phi_{,i} = -3\left(\frac{\dot{a}}{a}\right)^2 (a^2\dot{\delta u_i})$$

or simply[25]

$$(a^2\dot{\delta u_i}) = -\Phi_{,i}. \qquad \text{(C.9.39)}$$

Here we see an explicit demonstration of the assertion made above, in (C.9.32), that δu_i must be the gradient of some scalar field, U, which we can call the "velocity potential."

Writing $\delta u_i = U_{,i}$ in (C.9.39), taking the Fourier transform and dividing by ik gives

$$(a^2\dot{U}_k) = -\Phi_k. \qquad \text{(C.9.40)}$$

Meanwhile, doing the same in (C.9.36b) gives

$$\dot{\Phi}_k + H\Phi_k = -\frac{3}{2}H^2a^2U_k. \qquad \text{(C.9.41)}$$

Using (C.9.40) and its derivative with respect to time as substitutes for Φ_k and $\dot{\Phi}_k$ in (C.9.41) gives us an evolution equation for the velocity potential,

$$(a^2\ddot{U}_k) + H(a^2\dot{U}_k) = \frac{3}{2}H^2(a^2U_k). \qquad \text{(C.9.42)}$$

Using (C.9.16) to re-express this with respect to conformal time, and then (C.9.34b) to eliminate $\mathcal{H}$ in favour of ρ_b, gives a particularly simple expression of the evolution equation,

---

[25] Agrees with [72, equation (4.31)] when his pressure is set to 0.

$$(a^2U_k)'' = 4\pi\rho_b(a^2U_k). \tag{C.9.43}$$

The conclusion from (C.9.37) is that in a flat, pressureless, universe Φ is very closely approximated by a function that is independent of time. In this case, setting $\dot{\Phi} \approx 0$ in (C.9.36b) yields

$$a\delta u_i \approx -\frac{2}{3}\frac{\Phi_{,i}}{aH}. \tag{C.9.44}$$

This approximate relationship between the velocity and metric perturbations in a flat, dust-filled, universe is used in the collapse calculations.

Density Contrast

From the components of the Einstein equation in (C.9.36) we can also obtain the relationship between the density contrast and the metric perturbation. Dividing (C.9.36a) by $\frac{3}{2}\left(\frac{\dot{a}}{a}\right)^2$ and using (C.9.34b) gives[26]

$$\delta = \frac{2}{3}\left(\frac{1}{aH}\right)^2\frac{1}{H^2}\nabla^2\Phi - 2\frac{1}{H}\left(\dot{\Phi} + \frac{\dot{a}}{a}\Phi\right). \tag{C.9.45}$$

Transforming to frequency space and using (C.9.41), this can be written as

$$\delta_k = -\frac{2}{3}\left(\frac{k}{aH}\right)^2\Phi_k + 3Ha^2U_k. \tag{C.9.46}$$

Alternatively, using (C.9.37) to set $\dot{\Phi} = 0$ in (C.9.45) yields

$$\delta = \frac{2}{3}\left(\frac{1}{aH}\right)^2\frac{1}{H^2}\nabla^2\Phi - 2\Phi,$$

or in frequency space,

$$\delta_k = -\frac{2}{3}\left(\frac{k}{aH}\right)^2\Phi_k - 2\Phi_k. \tag{C.9.47}$$

C.9.3 Spherical Co-ordinates

The analysis of the Newtonian dust and inflaton perturbations carried out above was done in the context of a Cartesian co-ordinate grid on each surface of constant t. If one wishes to study spherically symmetric perturbations it is advantageous to express the results in spherical polar co-ordinates. Thanks to the simplicity of the results, doing this is fairly straight-forward. The relationship between the spatial Cartesian and polar co-ordinate systems is

---

[26] Agrees with [69, equation (5.28)].

$$x = r\sin\theta\cos\phi \tag{C.9.48a}$$
$$y = r\sin\theta\sin\phi \tag{C.9.48b}$$
$$z = r\cos\theta. \tag{C.9.48c}$$

The time co-ordinate is unchanged and remains orthogonal to all spatial co-ordinates so G_{tt}, given by (C.9.3a), will be unaffected. Since the spatial part of $G_{\mu\nu}$ is diagonal and the new co-ordinate system is orthogonal, the spatial part will remain diagonal and its components can be found as follows. For $i, j \in \{r, \theta, \phi\}$, we find

$$\begin{aligned} G_{ij} &= \left[G_{xx}\left(\frac{\partial x}{\partial i}\right)^2 + G_{yy}\left(\frac{\partial y}{\partial i}\right)^2 + G_{zz}\left(\frac{\partial z}{\partial i}\right)^2\right]\delta_{ij} \\ &= G_{xx}\left[\left(\frac{\partial x}{\partial i}\right)^2 + \left(\frac{\partial y}{\partial i}\right)^2 + \left(\frac{\partial z}{\partial i}\right)^2\right]\delta_{ij} \\ &= \ddot{\Phi} + 4\frac{\dot{a}}{a}\dot{\Phi} - \frac{1}{2}(1-4\Phi)\left(2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2\right)\begin{pmatrix}1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2\sin^2\theta\end{pmatrix}. \end{aligned} \tag{C.9.49}$$

The time-space part of $G_{\mu\nu}$ can be obtained from (C.9.3b) by (recalling that the time co-ordinate is both unchanged and orthogonal to all spatial co-ordinates)

$$G_{ti} = G_{tx}\frac{\partial x}{\partial i} + G_{ty}\frac{\partial y}{\partial i} + G_{tz}\frac{\partial z}{\partial i} = \left[\vec{\nabla}\left(2\dot{\Phi} + 2\frac{\dot{a}}{a}\Phi\right)\right]\cdot\left[\frac{\partial x}{\partial i}\hat{x} + \frac{\partial y}{\partial i}\hat{y} + \frac{\partial z}{\partial i}\hat{z}\right]$$

and spherical symmetry dictates that only G_{tr} will be non-zero which is simply

$$G_{tr} = 2\dot{\Phi}_{,r} + 2\frac{\dot{a}}{a}\Phi_{,r}. \tag{C.9.50}$$

C.10 The Derivation of (4.4.7)

The derivation begins with the integral expression for the second solution of a linear second-order differential equation which can be found in [3, equation 8.128]

$$\begin{aligned} u_{2k} &= u_{1k}\int^{\eta}\frac{1}{u_{1k}^2}\,\mathrm{d}\eta' \\ &= \frac{a'}{a^2\phi'}\int^{\eta}\left(\frac{a^2\phi'}{a'}\right)^2\mathrm{d}\eta' \\ &= \frac{\dot{a}}{a^2\dot{\phi}}\int^{t}\left(\frac{a^2\dot{\phi}}{\dot{a}}\right)^2\frac{1}{a}\,\mathrm{d}t' \end{aligned}$$

$$
\begin{aligned}
&= \frac{\dot{a}}{a^2\dot{\phi}} \int^t a \left(\frac{\dot{a}}{a}\right)^{-2} \dot{\phi}^2 \,\mathrm{d}t' \\
&= -4\pi \frac{\dot{a}}{a^2\dot{\phi}} \int^t a \left(\frac{\dot{a}}{a}\right)^{-2} \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\dot{a}}{a}\right) \mathrm{d}t' \\
&= 4\pi \frac{\dot{a}}{a^2\dot{\phi}} \int^t a \,\mathrm{d}\left(\frac{a}{\dot{a}}\right) \\
&= 4\pi \frac{\dot{a}}{a^2\dot{\phi}} \left(\frac{a^2}{\dot{a}} - \int^t \frac{a}{\dot{a}} \,\mathrm{d}a\right) \\
&= 4\pi \left(\frac{1}{\dot{\phi}} - \frac{\dot{a}}{a^2\dot{\phi}} \int^t \frac{a}{\dot{a}} \frac{\mathrm{d}a}{\mathrm{d}t} \,\mathrm{d}t'\right) \\
u_{2k} &= 4\pi \left(\frac{1}{\dot{\phi}} - \frac{\dot{a}}{a^2\dot{\phi}} \int^t a \,\mathrm{d}t'\right)
\end{aligned}
$$

where (C.9.6a) has been used.

C.11 Notes on Hypergeometric Functions

The hypergeometric function is defined by

$$
\mathrm{F}(\alpha,\beta;\gamma;z) = 1 + \frac{\alpha\beta}{\gamma\cdot 1} z + \frac{\alpha(\alpha+1)\beta(\beta+1)}{\gamma(\gamma+1)\cdot 1\cdot 2} z^2 + \frac{\alpha(\alpha+1)(\alpha+2)\beta(\beta+1)(\beta+2)}{\gamma(\gamma+1)(\gamma+2)\cdot 1\cdot 2\cdot 3} z^3 + \cdots . \tag{C.11.1}
$$

The sequence terminates if α or β is a negative integer or zero. For $\gamma = -n$, $n \in \{0,1,2,\dots\}$, the series is indeterminate if neither α nor β is equal to $-m$ where $m < n$ and m is a natural number.[27]

The generalized hypergeometric function is defined by

$$
{}_p\mathrm{F}_q(\alpha_1,\alpha_2,\dots,\alpha_p;\beta_1,\beta_2,\dots,\beta_q;z) = \sum_{k=0}^{\infty} \frac{(\alpha_1)_k(\alpha_2)_k\cdots(\alpha_p)_k}{(\beta_1)_k(\beta_2)_k\cdots(\beta_q)_k} \frac{z^k}{k!} \tag{C.11.2}
$$

where $(x)_n = 1\cdot x\cdots(x+n-1)$. The generalized hypergeometric function ${}_2\mathrm{F}_1$ is equivalent to the hypergeometric function F. The generalized hypergeometric function ${}_1\mathrm{F}_1$ is also called the confluent hypergeometric function, Kummer confluent hypergeometric function, Kummer function, etc., and is sometimes written $\Phi(a,b,z)$ or $M(a,b,z)$. Many other special functions can be written as special cases of ${}_1\mathrm{F}_1$. Generalized hypergeometric functions are symmetric under permutations of the α_i and of the β_i. The following properties of generalized hypergeometric functions are used in this document; the first four follow immediately from the definition above.

$$
{}_p\mathrm{F}_q(\alpha_1,\dots,\alpha_p;\beta_1,\dots,\beta_{q-1},\alpha_p;z) = {}_{p-1}\mathrm{F}_{q-1}(\alpha_1,\dots,\alpha_{p-1};\beta_1,\dots,\beta_{q-1};z) \tag{C.11.3}
$$

---

[27] See [28, equation (9.100)] for the definition. These properties can be found in the vicinity of the definition.

$$ {}_1\mathrm{F}_1(\alpha_1;\alpha_1;z) = \mathrm{e}^z \tag{C.11.4}$$

$$ {}_p\mathrm{F}_q(\alpha_1,\ldots,\alpha_{p-1},0;\beta_1,\ldots,\beta_q;z) = 1 \tag{C.11.5}$$

$$ {}_p\mathrm{F}_q(\alpha_1,\ldots,\alpha_p;\beta_1,\ldots,\beta_q;0) = 1 \tag{C.11.6}$$

[98, equation (07.20.03.0050.01)]:

$$ {}_1\mathrm{F}_1(1;\frac{1}{2};z) = \sqrt{\pi z}\,\mathrm{e}^z\,\mathrm{erf}(\sqrt{z}) + 1 \tag{C.11.7}$$

[98, equation (07.20.16.0001.01)]:

$$ {}_1\mathrm{F}_1(\alpha_1;\beta_1;z) = \mathrm{e}^z\,{}_1\mathrm{F}_1(\beta_1-\alpha_1;\beta_1;-z) \tag{C.11.8}$$

[98, equation (07.25.03.0006.01)]:

$$ {}_2\mathrm{F}_2(\alpha_1,1;\beta_1,2;z) = \frac{\beta_1-1}{(\alpha_1-1)z}\Big[{}_1\mathrm{F}_1(\alpha_1-1;\beta_1-1;z)-1\Big] \tag{C.11.9}$$

[98, equation (07.25.17.0010.01)]:

$$ \begin{aligned} &\alpha_2 z\,{}_2\mathrm{F}_2(\alpha_1+1,\alpha_2+1;\beta_1+1,\beta_2+1;z) + \\ &\qquad \beta_1\beta_2\Big[{}_2\mathrm{F}_2(\alpha_1,\alpha_2;\beta_1,\beta_2;z) - {}_2\mathrm{F}_2(\alpha_1+1,\alpha_2;\beta_1,\beta_2;z)\Big] = 0 \end{aligned} \tag{C.11.10}$$

or, for $\alpha_1 = 0$,

$$ {}_2\mathrm{F}_2(1,\alpha_2;\beta_1,\beta_2;z) = \frac{(\beta_1-1)(\beta_2-1)}{(\alpha_2-1)z}\Big[{}_2\mathrm{F}_2(1,\alpha_2-1;\beta_1-1,\beta_2-1;z)-1\Big] \tag{C.11.11}$$

[98, equation (07.25.17.0006.01)]:

$$ \begin{aligned} &-\alpha_1\,{}_2\mathrm{F}_2(\alpha_1+1,\alpha_2;\beta_1+1,\beta_2;z) + \\ &\qquad (\alpha_1-\beta_1)\,{}_2\mathrm{F}_2(\alpha_1,\alpha_2;\beta_1+1,\beta_2;z) + \beta_1\,{}_2\mathrm{F}_2(\alpha_1,\alpha_2;\beta_1,\beta_2;z) = 0 \end{aligned} \tag{C.11.12}$$

For ${}_2\mathrm{F}_2(\alpha_1,\alpha_2;\beta_1,\beta_2;z)$ where $\alpha_1,\beta_1 \in$ Naturals, repeated application of (C.11.12) will reduce α_1 to 1; repeated application of (C.11.11) will then reduce β_1 to 2, at which point (C.11.9) can be used to express the result in terms of ${}_1\mathrm{F}_1$. For example, starting from (C.11.12) with $\alpha_1 = 1$ and $\beta_1 = 3$,

$$ {}_2\mathrm{F}_2(2,\alpha_2;4,\beta_2;z) = 3\,{}_2\mathrm{F}_2(1,\alpha_2;3,\beta_2;z) - 2\,{}_2\mathrm{F}_2(1,\alpha_2;4,\beta_2;z). $$

Using (C.11.11) makes this

$$= 3\frac{2(\beta_2-1)}{(\alpha_2-1)z}\Big[{}_2\mathrm{F}_2(1,\alpha_2-1;2,\beta_2-1;z)-1\Big] - 2\frac{3(\beta_2-1)}{(\alpha_2-1)z}\Big[{}_2\mathrm{F}_2(1,\alpha_2-1;3,\beta_2-1;z)-1\Big]. \tag{C.11.13}$$

Using (C.11.9) on the first term and (C.11.11) again on the second terms leaves

$$\begin{aligned} &= 6\frac{(\beta_2-1)}{(\alpha_2-1)z}\left\{\frac{\beta_2-2}{(\alpha_2-2)z}\Big[{}_1\mathrm{F}_1(\alpha_2-2;\beta_2-2;z)-1\Big] - \frac{2(\beta_2-2)}{(\alpha_2-2)z}\Big[{}_2\mathrm{F}_2(1,\alpha_2-2;2,\beta_2-2;z)-1\Big]\right\} \\ &= 6\frac{(\beta_2-1)(\beta_2-2)}{(\alpha_2-1)(\alpha_2-2)z^2}\Big\{{}_1\mathrm{F}_1(\alpha_2-2;\beta_2-2;z)+1- 2\,{}_2\mathrm{F}_2(1,\alpha_2-2;2,\beta_2-2;z)\Big\}. \end{aligned}$$

Now using (C.11.9) on the second terms gives

$${}_2\mathrm{F}_2(2,\alpha_2;4,\beta_2;z) = 6\frac{(\beta_2-1)(\beta_2-2)}{(\alpha_2-1)(\alpha_2-2)z^2}\left\{{}_1\mathrm{F}_1(\alpha_2-2;\beta_2-2;z)+1- \frac{2(\beta_2-3)}{(\alpha_2-3)z}\Big[{}_1\mathrm{F}_1(\alpha_2-3;\beta_2-3;z)-1\Big]\right\}. \tag{C.11.14}$$

The following is useful in finding the large-z asymptotic limit of the confluent hypergeometric function.[28]

$${}_1\mathrm{F}_1(a;b;z) = \frac{\Gamma(b)}{\Gamma(b-a)}e^{i\pi a}z^{-a}\left[\sum_{n=0}^{R-1}\frac{(a)_n(1+a-b)_n}{n!}(-z)^{-n} + \mathrm{O}(|z|^{-R})\right] + \frac{\Gamma(b)}{\Gamma(a)}e^{z}z^{a-b}\left[\sum_{n=0}^{S-1}\frac{(b-a)_n(1-a)_n}{n!}z^{-n} + \mathrm{O}(|z|^{-S})\right]. \tag{C.11.15}$$

---

[28] See [2, Section 13.5.1].

C.12 The Derivation of (5.2.14)

Writing out the spherical Bessel function as $\mathrm{j}_1(z) = \sqrt{\frac{\pi}{2z}}\,\mathrm{J}_{3/2}(z)$, the integral can be written in terms of the generalized hypergeometric function, ${}_3\mathrm{F}_3$. Using[29]

$$\begin{aligned}\int_0^\infty & x^{\lambda-1}\mathrm{e}^{-\alpha x^2}\,\mathrm{J}_\mu(\beta x)\,\mathrm{J}_\nu(\beta x)\,\mathrm{d}x \\ &= 2^{-\nu-\mu-1}\,\alpha^{-\frac{1}{2}(\nu+\lambda+\mu)}\,\beta^{\nu+\mu}\,\frac{\Gamma(\frac{1}{2}\lambda+\frac{1}{2}\mu+\frac{1}{2}\nu)}{\Gamma(\mu+1)\Gamma(\nu+1)}\times \\ &\quad {}_3\mathrm{F}_3\left[\frac{\nu}{2}+\frac{\mu}{2}+\frac{1}{2},\frac{\nu}{2}+\frac{\mu}{2}+1,\frac{\nu+\mu+\lambda}{2};\mu+1,\nu+1,\mu+\nu+1;-\frac{\beta^2}{\alpha}\right]; \\ &\qquad \Re(\nu+\lambda+\mu)>0,\ \Re(\alpha)>0,\end{aligned}$$

with $\lambda = n_{s\Phi}$, $\alpha = \frac{r_0^2}{r^2}$, $\beta = 1$, and $\mu = \nu = \frac{3}{2}$ gives

$$\left\langle\left(\frac{\Phi'}{aH}\right)^2\right\rangle = \frac{\sigma^2}{18}\Gamma\Big(\frac{n_{s\Phi}+3}{2}\Big)\left(\frac{1}{raH}\right)^{n_{s\Phi}+1}\left(\frac{r}{r_0}\right)^{n_{s\Phi}+3}\times \\ {}_2\mathrm{F}_2\left[2,\frac{n_{s\Phi}+3}{2};\frac{5}{2},4;-\left(\frac{r}{r_0}\right)^2\right] \qquad \text{(C.12.1)}$$

for $n_{s\Phi} > -3$. Using the relationships in Appendix C.11, this can be written in terms of lower-ordered hypergeometric functions. Using (C.11.14) we have

$${}_2\mathrm{F}_2(2,\frac{n_{s\Phi}+3}{2};4,\frac{5}{2};-z^2) = \frac{18}{(n_{s\Phi}+1)(n_{s\Phi}-1)z^4}\Bigg\{{}_1\mathrm{F}_1(\frac{n_{s\Phi}-1}{2};\frac{1}{2};-z^2)+1- \\ \frac{2}{(n_{s\Phi}-3)z^2}\Big[{}_1\mathrm{F}_1(\frac{n_{s\Phi}-3}{2};-\frac{1}{2};-z^2)-1\Big]\Bigg\}.$$

Although the right-hand side is formally equivalent to ${}_2\mathrm{F}_2(2,\frac{n_{s\Phi}+3}{2};4,\frac{5}{2};-z^2)$, it's no longer clear that it's well defined for all $n_{s\Phi} > -3$. In particular we have problems at $n_{s\Phi} = -1, 1, 3$. This is simply because some of the identities used to simplify (C.12.1) reduce to $0 = 0$ at these spectral indices. Let's investigate the behaviour at $n_{s\Phi} = 1$ more carefully, and ignore the problems at the other two spectral indices since they are outside the range of physically relevant values. For $n_{s\Phi} = 1$, it is the application of (C.11.9) and (C.11.11) following (C.11.13) that gives us problems so let's stop there and write that

$${}_2\mathrm{F}_2(2,2;4,\frac{5}{2};-z^2) = \frac{9}{z^2}\Bigg\{{}_2\mathrm{F}_2(1,1;\frac{3}{2},3;-z^2) - {}_2\mathrm{F}_2(1,1;\frac{3}{2},2;-z^2)\Bigg\}.$$

Altogether, then,

---

[29] See [28, equation (6.633.5)].

$$\left\langle\left(\frac{\Phi'}{aH}\right)^2\right\rangle = \frac{\sigma^2}{4}\Gamma\Big(\frac{n_{\mathrm{s}\Phi}-1}{2}\Big)\left(\frac{1}{r_0aH}\right)^{n_{\mathrm{s}\Phi}-1}(raH)^{-2}\times$$
$$\left\{{}_1\mathrm{F}_1\left(\frac{n_{\mathrm{s}\Phi}-1}{2};\frac{1}{2};-\Big(\frac{r}{r_0}\Big)^2\right)+1-\right.$$
$$\left.\frac{2}{n_{\mathrm{s}\Phi}-3}\Big(\frac{r_0}{r}\Big)^2\left[{}_1\mathrm{F}_1\left(\frac{n_{\mathrm{s}\Phi}-3}{2};-\frac{1}{2};-\Big(\frac{r}{r_0}\Big)^2\right)-1\right]\right\} \tag{C.12.2a}$$

for $n_{\mathrm{s}\Phi} > -3$, $n_{\mathrm{s}\Phi} \neq -1, 1, 3$, and

$$\left\langle\left(\frac{\Phi'}{aH}\right)^2\right\rangle = \frac{\sigma^2}{2}\left(\frac{1}{r_0aH}\right)^2\left\{{}_2\mathrm{F}_2\left(1,1;\frac{3}{2},3;-\Big(\frac{r}{r_0}\Big)^2\right) - {}_2\mathrm{F}_2\left(1,1;\frac{3}{2},2;-\Big(\frac{r}{r_0}\Big)^2\right)\right\} \tag{C.12.2b}$$

for $n_{\mathrm{s}\Phi} = 1$.

Asymptotic Behaviour

The large-r asymptotic behaviour is obtained from (C.12.2a) and (C.11.15). From (C.11.15), the leading-order term at large z in ${}_1\mathrm{F}_1(a;b;-z^2)$ is

$$\begin{aligned}{}_1\mathrm{F}_1(a;b;-z^2) &\approx \frac{\Gamma(b)}{\Gamma(b-a)}\mathrm{e}^{\mathrm{i}\pi a}(-1)^{-a}z^{-2a}\\ &= \frac{\Gamma(b)}{\Gamma(b-a)}z^{-2a}\end{aligned} \tag{C.12.3}$$

so at large r,

$$\begin{aligned}&{}_1\mathrm{F}_1\left(\frac{n_{\mathrm{s}\Phi}-1}{2};\frac{1}{2};-\Big(\frac{r}{r_0}\Big)^2\right)+1-\frac{2}{n_{\mathrm{s}\Phi}-3}\Big(\frac{r_0}{r}\Big)^2\left[{}_1\mathrm{F}_1\left(\frac{n_{\mathrm{s}\Phi}-3}{2};-\frac{1}{2};-\Big(\frac{r}{r_0}\Big)^2\right)-1\right]\\ &\quad\approx \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{1}{2}-\frac{n_{\mathrm{s}\Phi}-1}{2})}\Big(\frac{r}{r_0}\Big)^{-(n_{\mathrm{s}\Phi}-1)}+1+\\ &\qquad\qquad\frac{2}{3-n_{\mathrm{s}\Phi}}\Big(\frac{r}{r_0}\Big)^{-2}\left[\frac{\Gamma(-\frac{1}{2})}{\Gamma(-\frac{1}{2}-\frac{n_{\mathrm{s}\Phi}-3}{2})}\Big(\frac{r}{r_0}\Big)^{-(n_{\mathrm{s}\Phi}-3)}-1\right]\\ &\quad=\left[\frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{2-n_{\mathrm{s}\Phi}}{2})}+\frac{2}{3-n_{\mathrm{s}\Phi}}\frac{\Gamma(-\frac{1}{2})}{\Gamma(\frac{2-n_{\mathrm{s}\Phi}}{2})}\right]\Big(\frac{r}{r_0}\Big)^{1-n_{\mathrm{s}\Phi}}+1+\frac{2}{3-n_{\mathrm{s}\Phi}}\Big(\frac{r}{r_0}\Big)^{-2}\end{aligned}$$

which, for $n_{\mathrm{s}\Phi} \sim 1$, is

$$\approx \frac{2}{3-n_{\mathrm{s}\Phi}}\Big(\frac{r}{r_0}\Big)^{-2}.$$

From this, at large r and $n_{s\Phi} \sim 1$,

$$\left\langle \left(\frac{\Phi'}{aH}\right)^2 \right\rangle \approx \frac{\sigma^2}{2(3-n_{s\Phi})}\Gamma\Big(\frac{n_{s\Phi}-1}{2}\Big)\left(\frac{1}{r_0 aH}\right)^{n_{s\Phi}-3}(raH)^{-4}. \tag{C.12.4}$$

C.13 The Derivation of (5.4.10)

Starting from (5.4.9),

$$\left\langle \tau^{-\frac{2}{3}}(r_1)\tau^{-\frac{2}{3}}(r_2)\right\rangle = \frac{1}{r_1 aH}\left(\frac{1}{raH}\right)^{n_{s\Phi}+2}\int x^{n_{s\Phi}}\,\mathrm{j}_1(\frac{r_1}{r}x)\,\mathrm{j}_1(x)\exp\left[-\frac{r_0^2}{r^2}x^2\right]\mathrm{d}x.$$

Writing the spherical Bessel functions as $\mathrm{j}_1(z) = \sqrt{\frac{\pi}{2z}}\,\mathrm{J}_{\frac{3}{2}}(z)$, the integral becomes

$$\frac{\pi}{2}\sqrt{\frac{r}{r_1}}\int x^{n_{s\Phi}-1}\,\mathrm{J}_{\frac{3}{2}}(\frac{r_1}{r}x)\,\mathrm{J}_{\frac{3}{2}}(x)\exp\left[-\frac{r_0^2}{r^2}x^2\right]\mathrm{d}x.$$

Using[30]

$$\begin{aligned}&\int_0^\infty x^{\lambda+1}e^{-\alpha x^2}\,\mathrm{J}_\mu(\beta x)\,\mathrm{J}_\nu(\gamma x)\,\mathrm{d}x\\
&\qquad= \frac{\beta^\mu\gamma^\nu\alpha^{-\frac{1}{2}(\mu+\nu+\lambda+2)}}{2^{\nu+\mu+1}\Gamma(\nu+1)}\times\\
&\qquad\qquad\sum_{m=0}^{\infty}\frac{\Gamma(m+\frac{1}{2}\nu+\frac{1}{2}\mu+\frac{1}{2}\lambda+1)}{m!\Gamma(m+\mu+1)}\left(-\frac{\beta^2}{4\alpha}\right)^m \mathrm{F}(-m,-\mu-m;\nu+1;\frac{\gamma^2}{\beta^2}),\\
&\qquad\qquad\qquad\Re(\alpha)>0,\ \Re(\mu+\nu+\lambda)>-2,\ \beta>0,\ \gamma>0,\end{aligned}$$

with $\lambda = n_{s\Phi}-2$, $\alpha = r_0^2/r^2$, $\mu=\nu=3/2$, $\beta = r_1/r$, and $\gamma = 1$, we get

$$\begin{aligned}\left\langle \tau^{-\frac{2}{3}}(r_1)\tau^{-\frac{2}{3}}(r_2)\right\rangle &= \frac{\pi}{2}\frac{1}{(r_1 aH)^{3/2}}\left(\frac{1}{raH}\right)^{n_{s\Phi}+\frac{3}{2}}\frac{(\frac{r_1}{r})^{\frac{3}{2}}(\frac{r}{r_0})^{(n_{s\Phi}+3)}}{2^4\Gamma(\frac{5}{2})}\times\\
&\qquad\sum_{m=0}^{\infty}\frac{\Gamma(m+\frac{3}{2}+\frac{1}{2}n_{s\Phi})}{m!\,\Gamma(m+\frac{5}{2})}\left(-\frac{r_1^2}{4r_0^2}\right)^m \mathrm{F}\left(-m,-\frac{3}{2}-m;\frac{5}{2};\frac{r^2}{r_1^2}\right)\\
&= \frac{\sqrt{\pi}}{24}\left(\frac{1}{r_0 aH}\right)^{n_{s\Phi}+3}\times\\
&\qquad\sum_{m=0}^{\infty}\frac{\Gamma(m+\frac{1}{2}(n_{s\Phi}+3))}{m!\,\Gamma(m+\frac{5}{2})}\left(-\frac{r_1^2}{4r_0^2}\right)^m \mathrm{F}\left(-m,-\frac{3}{2}-m;\frac{5}{2};\frac{r^2}{r_1^2}\right)\end{aligned}$$

for $n_{s\Phi} > -3$.

---

[30] See [28, equation (6.633.1)].

Appendix D

Properties of Distributions and Random Variables

D.1 Propagation of Distribution Densities

We must first look at the way a distribution density propagates through a function. Assume that $y = f(x)$ gives a relationship between values of x and values of y where f is a continuous function whose first derivative exists. If values of x are chosen at random according to a distribution density, $P(x)$, then the probability that a particular choice will lie between x and $x + \mathrm{d}x$ is $P(x)\,\mathrm{d}x$. Furthermore, let the probability that a value of y is observed between y and $y + \mathrm{d}y$ be $P(y)\,\mathrm{d}y$ where $P(y)$ is the distribution density for y (the two densities being distinguished by their parameters). Of course $P(y)\,\mathrm{d}y = P(x)\,\mathrm{d}x$ where $y = f(x)$ and $\mathrm{d}y$ is the interval at y corresponding to the interval $\mathrm{d}x$ at x. So

$$P(y) = P(x)\frac{\mathrm{d}x}{\mathrm{d}y} = P(x)\left|f'\right|^{-1} = \sum_i \frac{P(x_i)}{|f'(x_i)|} \tag{D.1.1}$$

where in the last step the possibility that several choices of x may result in the same value of y has been taken into consideration so that $\{x_i\}$ is the set of solutions to $x = f^{-1}(y)$. In particular, if

$$y = \alpha x^n$$

then the solution of $x = f^{-1}(y)$ is

$$x = \left(\frac{y}{\alpha}\right)^{\frac{1}{n}}$$

and the slope of f is

$$\frac{\mathrm{d}y}{\mathrm{d}x} = n\alpha x^{n-1}$$

which, when evaluated at the solution above, is

$$\left.\frac{\mathrm{d}y}{\mathrm{d}x}\right|_{x=f^{-1}(y)} = n\alpha^{\frac{1}{n}} y^{1-\frac{1}{n}}.$$

Altogether, then,

$$P(y) = \left|\frac{1}{n}\alpha^{-\frac{1}{n}} y^{\frac{1}{n}-1}\right| P\left[x = \left(\frac{y}{\alpha}\right)^{\frac{1}{n}}\right]. \tag{D.1.2}$$

Let's generalize this technique to functions of more than one variable. Assume that we have a quantity, ϕ, whose value is defined everywhere in an n-dimensional volume so that $\phi = f(\vec{x})$ where f is continuous and has a gradient everywhere. Again, the problem is to find a distribution density for ϕ, $P(\phi)$, corresponding to choices of $\vec{x}$ made according to a distribution density $P(\vec{x})$ (the two densities, again, being distinguished by their parameters). The point to notice here is that when given a ϕ, the solution of $\vec{x} = f^{-1}(\phi)$ is not a single point as in the 1-dimension case but a surface of dimension $n-1$. The probability that a value of ϕ occurs between ϕ and $\phi + \mathrm{d}\phi$ is, therefore,

$$P(\phi)\,\mathrm{d}\phi = \int_\sigma [P(\vec{x})\,\mathrm{d}n]\,\mathrm{d}A$$

where σ is the surface of constant ϕ, $\mathrm{d}A$ is the area element on the surface and $\mathrm{d}n$ is the interval at $\vec{x}$, perpendicular to the surface, which corresponds to the interval $\mathrm{d}\phi$. Continuing

$$\begin{aligned}
P(\phi)\,\mathrm{d}\phi &= \int_\sigma [P(\vec{x})\,\mathrm{d}n]\,\mathrm{d}A \\
P(\phi) &= \int_\sigma P(\vec{x})\frac{\mathrm{d}n}{\mathrm{d}\phi}\,\mathrm{d}A \\
&= \int_\sigma P(\vec{x})\left(\frac{\mathrm{d}\phi}{\mathrm{d}n}\right)^{-1}\mathrm{d}A \\
P(\phi) &= \int_\sigma \frac{P(\vec{x})}{\left|\vec{\nabla} f(\vec{x})\right|}\,\mathrm{d}A.
\end{aligned}$$

If there are multiple, disjoint, surfaces of the same constant ϕ then their contributions must be added

$$P(\phi) = \sum_i \int_{\sigma_i} \frac{P(\vec{x})}{\left|\vec{\nabla} f(\vec{x})\right|}\,\mathrm{d}A_i \tag{D.1.3}$$

where $\{\sigma_i\}$ is the set of surfaces for which $f\left(\vec{x} \in \sigma_i\right) = \phi$. To illustrate the use of this method for mapping distributions, two simple examples are presented in Appendix D.2.

D.2 Gaussian Random Variables and Processes

The first two computations are standard results for the properties of Gaussian random variables. They are presented here in the manner of exercises to illustrate the use of the techniques from the previous section. This will hopefully make some of the other applications of these techniques throughout this document more transparent.

If x and y are two independent normally distributed random variables so that

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \mathrm{e}^{-\frac{1}{2\sigma_x^2}x^2}$$
$$P(y) = \frac{1}{\sqrt{2\pi}\sigma_y} \mathrm{e}^{-\frac{1}{2\sigma_y^2}y^2}$$

then the joint probability, $P(x,y)$, is

$$P(x,y) = P(x)P(y).$$

Using the above, one gets the following.

D.2.1 The Distribution of $z = ax$

where a is a constant. This is a simple problem that serves as a good introduction to the technique. We have a one-dimensional space with co-ordinate x in which is defined the function $z(x) = ax$. A surface of constant z in this space is the point given by

$$x = \frac{1}{a}z.$$

The magnitude of the gradient of $z(x)$ is a constant and is

$$\left|\frac{\mathrm{d}z}{\mathrm{d}x}\right| = |a|\,.$$

Using these in (D.1.1) gives

$$\begin{aligned} P(z) &= \frac{P(x = \frac{1}{a}z)}{|a|} \\ &= \frac{1}{\sqrt{2\pi}(a\sigma_x)} \exp\left[-\frac{1}{2}\frac{z^2}{(a\sigma_x)^2}\right] \\ &= \frac{1}{\sqrt{2\pi}\sigma_z} \exp\left[-\frac{1}{2}\frac{z^2}{\sigma_z^2}\right] \end{aligned}$$

where

$$\sigma_z = a\sigma_x.$$

So $z = ax$ is a Gaussian random variable with standard deviation $\sigma_z = a\sigma_x$ which we know as the standard rule for propagating uncertainty through a linear co-efficient.[1]

D.2.2 The Distribution of $z = x + y$

This is a slightly more complex example. In this case we have a two-dimensional space with co-ordinates x and y in which is defined the function $z(x, y) = x + y$. Starting by identifying surfaces of constant z in this space, let's express these as a constraint on y:

$$y = z - x. \tag{D.2.1}$$

The magnitude of the gradient of z is a constant, namely

$$\left|\vec{\nabla} z\right| = \left|\frac{\partial z}{\partial x}\hat{x} + \frac{\partial z}{\partial y}\hat{y}\right| = |\hat{x} + \hat{y}| = \sqrt{2}.$$

The area element on a surface of constant z is obtained using (D.2.1) and is the line element

$$\mathrm{d}l = \sqrt{(\mathrm{d}x)^2 + (\mathrm{d}y)^2} = \mathrm{d}x\sqrt{1 + \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2} = \sqrt{2}\,\mathrm{d}x.$$

Substituting the above into (D.1.3),

$$\begin{aligned} P(z) &= \int_{x=-\infty}^{+\infty} \frac{P(x)P(y = z - x)}{\sqrt{2}} \sqrt{2}\,\mathrm{d}x \\ &= \frac{1}{2\pi\sigma_x\sigma_y} \int_{-\infty}^{+\infty} \exp\left[-\frac{1}{2}\frac{x^2}{\sigma_x^2} - \frac{1}{2}\frac{(z-x)^2}{\sigma_y^2}\right] \mathrm{d}x \\ &= \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\frac{z^2}{\sigma_x^2 + \sigma_y^2}\right] \int_{-\infty}^{+\infty} \exp\left[-\frac{1}{2}\frac{[(\sigma_x^2 + \sigma_y^2)x - \sigma_x^2 z]^2}{\sigma_x^2\sigma_y^2(\sigma_x^2 + \sigma_y^2)}\right] \mathrm{d}x \end{aligned}$$

let

$$\begin{aligned} t &= (\sigma_x^2 + \sigma_y^2)x - \sigma_x^2 z \\ \mathrm{d}x &= \frac{1}{\sigma_x^2 + \sigma_y^2}\,\mathrm{d}t \end{aligned}$$

so

$$\begin{aligned} P(z) &= \frac{1}{2\pi\sigma_x\sigma_y(\sigma_x^2 + \sigma_y^2)} \exp\left[-\frac{1}{2}\frac{z^2}{\sigma_x^2 + \sigma_y^2}\right] \int_{-\infty}^{+\infty} \exp\left[-\frac{1}{2}\frac{t^2}{\sigma_x^2\sigma_y^2(\sigma_x^2 + \sigma_y^2)}\right] \mathrm{d}t \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_x^2 + \sigma_y^2}} \exp\left[-\frac{1}{2}\frac{z^2}{\sigma_x^2 + \sigma_y^2}\right]. \end{aligned}$$

---

[1] See [88].

Therefore, z is a normally distributed random variable with a standard deviation of

$$\sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2}.$$

This is the standard result for Gaussian random variables often described as "adding widths in quadrature."[2]

D.3 The Dirac δ

The following definitions of the Dirac δ are found in [40].

$$\nabla^2 \frac{1}{|\vec{x}-\vec{x}'|} = -4\pi\delta(\vec{x}-\vec{x}') \tag{D.3.1}$$

$$\delta(x-x') = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \mathrm{e}^{ik(x-x')}\,\mathrm{d}k \tag{D.3.2}$$

$$\frac{1}{k}\delta(k'-k) = \int_0^\infty x\,\mathrm{J}_m(kx)\,\mathrm{J}_m(k'x)\,\mathrm{d}x \tag{D.3.3}$$

$$\frac{\pi}{2k^2}\delta(k-k') = \int_0^\infty r^2\,\mathrm{j}_l(kr)\,\mathrm{j}_l(k'r)\,\mathrm{d}r \tag{D.3.4}$$

$$\delta(z-z') = \frac{1}{\pi}\int_0^\infty \mathrm{d}k\cos[k(z-z')] \tag{D.3.5}$$

$$\delta(\phi-\phi') = \frac{1}{2\pi}\sum_{m=-\infty}^{+\infty} \mathrm{e}^{im(\phi-\phi')}. \tag{D.3.6}$$

See [40, (1.31)], [40, (2.47)], [40, (3.108)], and [40, (3.112)] for the first four respectively and [40, 3.139] for both of the last two. Note that (D.3.5) is simply (D.3.2) with the odd part of the integrand discarded.

The following definitions of the Dirac δ as the limit of sequences of functions can be found in [3]. The Dirac δ is

$$\delta(x) = \lim_{n\to\infty} \delta_n(x) \tag{D.3.7}$$

where $\delta_n(x)$ can be defined as

$$\delta_n(x) = \begin{cases} 0, & x < -\frac{1}{2n} \\ n, & -\frac{1}{2n} < x < +\frac{1}{2n} \\ 0, & x > +\frac{1}{2n} \end{cases} \tag{D.3.8}$$

$$\delta_n(x) = \frac{n}{\sqrt{\pi}}\exp\left(-n^2x^2\right) \tag{D.3.9}$$

$$\delta_n(x) = \frac{n}{\pi}\frac{1}{1+n^2x^2} \tag{D.3.10}$$

---

[2] See [88].

$$\delta_n(x) = \frac{\sin nx}{\pi x} = \frac{1}{2\pi}\int_{-n}^{+n} \mathrm{e}^{\mathrm{i}xt}\, \mathrm{d}t. \tag{D.3.11}$$

See [3, (1.171)] [3, (1.172)] [3, (1.173)] [3, (1.174)] respectively and note that (D.3.11) is equivalent to (D.3.2).

Some properties of the Dirac δ are

$$\delta(ax) = \frac{1}{a}\delta(x), \text{ for } a > 0 \tag{D.3.12}$$

$$\delta(g(x)) = \sum_{\{a|g(a)=0,g'(a)\neq 0\}} \frac{\delta(x-a)}{|g'(a)|}. \tag{D.3.13}$$

See [3, (1.179)], [3, (1.180)], respectively. Of course, the first is a special case of the second.